

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 22, 2002, 02:13:10 ; Search time 2452.42 Seconds

(without alignments)  
10947.858 Million cell updates/sec

Title: US-09-758-017A-1

Perfect score: 1283

Sequence: 1 gacatccgctgcaaatatg.....agaataatattatatta 1283

Scoring table: IDENTITY\_NUC

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database : GenBank.\*

1: gb\_ba.\*  
2: gb\_hcg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: em\_da.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vl.\*  
30: em\_hlg\_hum.\*  
31: em\_hlg\_inv.\*  
32: em\_hlg\_other.\*  
33: em\_hlg\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No. Score Match length DB ID Description

1	1278.2	99.6	1283	5	GM0275971	AJ275971	Gadus mor
2	1164.2	90.7	1355	5	GM0275972	AJ275972	Gadus mor
3	364.2	28.4	1857	10	MM055041	U55041	Mus musculu
4	362.6	28.3	2002	10	BC011039	BC011039	Mus muscu
5	361	28.1	1875	10	MM08975	X99018	M. musculus
6	361	28.1	1951	10	BC015205	X08975	M. musculus
7	355.4	27.7	2045	9	BC015205	BC015205	Homo sapi
8	355.4	27.7	2053	6	A64377	A64377	Sequence 1
9	355.4	27.7	2057	6	HS09008	Y09008	H. sapiens m
10	355.4	27.7	2062	9	HS09008	X15653	Human cDNA
11	355.4	27.7	2122	6	AX281824	AX281824	Sequence
12	354.9	27.7	1737	10	BC004037	BC004037	Mus muscu
13	262.8	20.5	189672	9	CNS01DXY	AL136059	Human chr
14	262.8	20.5	213353	2	AC010258	AC010258	Homo sapi
15	261.2	20.4	123159	2	AC091880	AC091880	Homo sapi
16	261.2	20.4	171125	9	AC025437	AC025437	Homo sapi
17	226.8	17.7	130932	9	AC092658	AC092658	Homo sapi
18	210.8	16.4	12337	1	AE005488	AE005488	Escherich
19	210	16.4	21722	1	AE008820	AE008820	Salmonell
20	210	16.4	274050	1	AL627276	AL627276	Salmonell
21	207.6	16.2	1532	1	ECOUNG	J03725	E. coli ung
22	207.6	16.2	1532	6	AX277028	AX277028	Sequence
23	207.6	16.2	10272	1	ECORBAB	D13169	E. coli gene
24	207.6	16.2	12510	1	AE000344	AE000344	Escherich
25	207.6	16.2	16446	1	D90886	D90886	E. coli geno
26	207.6	16.2	29254	1	ECCK12RIII	D64044	Escherichia
27	206.8	15.7	258050	1	AJ414153	AJ414153	Yersinia
28	201.4	15.5	270365	1	AP002562	AP002562	Escherich
29	199.2	15.5	12918	1	AE001926	AE001926	Deinococc
30	198.2	15.4	78736	2	AC095886	AC095886	Rattus no
31	186.2	14.5	10937	1	AE004306	AE004306	Vibrio ch
32	180	14.0	657	6	AX083978	AX083978	Sequence
33	180	14.0	11847	1	U32687	U32687	Haemophilu
34	178.6	13.9	158325	2	AC087564	AC087564	Homo sapi
35	172.2	13.4	10144	1	AE002470	AE002470	Homo sapi
36	172.2	13.4	191549	9	AC010582	AC010582	Neisseria
37	172.2	13.4	199420	9	CNS01DXY	AL136040	Human chr
38	172.2	13.4	349980	6	AX044031	AX044031	Sequence
39	172.2	13.4	349980	6	AX044032	AX044032	Sequence
40	170.8	13.3	42071	1	SC2661	AL339949	Streptomy
41	169	13.2	12736	1	AE006042	AE006042	Pasteurel
42	168.8	13.2	333800	1	SME591792	AL591792	Sinorhizo
43	165	12.9	12258	1	AE009193	AE009193	Agrobacte
44	165	12.9	13567	1	AE008159	AE008159	Agrobacte
45	161	12.5	331801	1	NMA422491	AL162755	Neisseria

#### ALIGNMENTS

RESULT 1  
LOCUS GM0275971 1283 bp mRNA linear VRT 04-APR-2000  
DEFINITION Gadus morhua mRNA for uracil-DNA glycosylase precursor (ung1 gene).  
ACCESSION AJ275971.1 GI:7413911  
VERSION AJ275971.1 GI:7413911  
KEYWORDS precursor; ung1 gene; uracil-DNA glycosylase.  
SOURCE Atlantic cod.  
ORGANISM Gadus morhua  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Gadus.  
REFERENCE  
Lanes, O. and Willassen, N.P.  
Identification, cloning and expression of nuclear and mitochondrial  
uracil-DNA glycosylase from Atlantic cod(Gadus morhua).  
Characterisation of the cold-active catalytic domain  
Unpublished  
JOURNAL  
REFERENCE  
Lanes, O.  
Direct Submission  
Submitted (02-MAR-2000) Lanes O., Department of Biotechnology,  
University of Tromsøe, Institute of Medical Biology, Medical

FEATURES		Facility, N-9037, NORWAY
source		Location/Qualifiers
		1..1283
		/organism="Gadus morhua"
		/db_xref="taxon:8049"
		/tissue_type="liver"
		/dev_stage="adult"
		/country="Norway"
		18..923
gene		/gene="ung1"
		18..923
CDS		/gene="ung1"
		/EC_number="3.2.2.3"
		/codon_start=1
		/product="uracil-DNA glycosylase precursor"
		/protein_id="CAB85707.1"
		/db_xref="gi:7413912"
		/translation="MLFKILGICORCISNRVLPGLLIPOTLCFSKLMKITPKLRSSN VEOKTSPOLSYEOLERAKNKKALDKIRAKAPAGGEMPRRLAEFEKPYKOL MSFVADERSRHVYPPADOVYSWTEMCIDQVKVYILGODPHGNQAHGICFSYQK VPPPSLVNIXKELCTDIDGFKHPGHGDLGSAWAKGVLLNLNVLTVRAHQASHKDRG WETFDVAIKMLSVNREGVLELLMGSYAHKKKQATIDRRKHVLAQVHPSPLSHRGFL GCKHPSKANGLIKLSGTEPIWMRAL"
BASE COUNT		364 a 279 c 280 g 360 t
ORIGIN		
Query Match		99.6%; Score 1278.2; DB 5; Length 1283;
Best Local Similarity		99.8%; Pred. No. 0;
Matches 1280; Conservative		0; Mismatches 3; Indels 0; Gaps 0;
QY	1	gacatcgccttgcaaatatgcttgctcaagttaggttaagccaagaatgcaatacctaa 60
DB	1	GACATCCGCTTCAAAATATGTTTCMAAGTTAGGGTTATGCCAAAGATCATATCATCA 60
QY	61	atcgggtgttaccagagttactcaatccccaacttatgttttcttaataataaga 120
DB	61	ATCGGGTCTTACCAGGTACTTACTAATTCCTCCCAACTTATGTTTCTTAATTATCAAG 120
QY	121	taacgcgaagaactgagtgctcctcaaatgttgaaacaaagacgtccacagcttt 180
DB	121	TAAAGCGCAAGAAACTGAGGTCCTCAATGTGGAACAAAGACGTATGCCACACTTT 180
QY	181	caagtggagcagcttgagaagaatgagccaaataagaagaagcagcgttgacaaga 240
DB	181	CAGTGGAGCAGCTGGAATAATGGCCAAATAAGAAACAGCGCTTGACAAATATAGAG 240
QY	241	caaaaagcagcgtgcaggttttcgagagacttgagagaagagagcgtgctgcagagt 300
DB	241	CAAAAGCAACGCTGCAGGTTTCGAGAGACTTGGAGAGAGAGCTGGCTGCAGAGTTTG 300
QY	301	aaaagccatacttcaaaacatltgatlgtccttctgttaactgataagagagccgtcaacgc 360
DB	301	AAAAGCATACTTCAAAACAATTGATGCTCTTGTACTGATGAGAGAGCGGTCAACCG 360
QY	361	tctaccaccagcctgatacaagtgtacagttcgacagagatgctgtgacatccaagtgtga 420
DB	361	TCCTACCCACCGGCTGATCAAGTGTACAGTTGGACAGATGCTGTGACATTCAGAGTGA 420
QY	421	aagtgatgattcttaggcgaagaccccttaccagagctgccaaagaagacatgagcttgtt 480
DB	421	AAGTACTGATTTCTAGGCGGAGACCTTACCACGAGTCCCAACCAAGACATGGACTGTGT 480
QY	481	tcaagtgtgaaaagcaggttcccccctcccccagctcgttgaaacatacaagaattgt 540
DB	481	TCAGTGTGCAAAAGCAGTTCCCGCCGCCCGACGTCTGTGAACAATATCAAGAATTGT 540
QY	541	gtaccgacaatgtatggcttcaagcatcctgtgacatgagatctaaacggatggcgaataac 600
DB	541	GTACCGACATTTGATGGCTTCAAGCATCTTGACATGAGATCATTAACGGATGGCAGAAAC 600
QY	601	aagggtgtgtgtcttaacgcggtgctgacggtgaggccatacaggccaactcccaaa 660
DB	601	AAGGGTGTGTGCTTAAACGGGCTGTGACCGTGGGGCCCATCAGGCCAACTCCCA 660
QY	661	aggacagagcctgtagaaccttcacagcgtgtgatacaagtgtgtagcgtcaaacggag 720
DB	661	AGGACAGAGGCTGGAGACCTTCACCGACGCTGTGATCAAGTGGCTGACGCTCAACCGG 720
QY	721	aaggatgcgttttccctgtgtgtggtgtcaacgccaataagaaggagagacatcgaca 780
DB	721	AAGGATGCGTTTCCCTGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 780
QY	781	ggaacgtcaccaatgctgtgcaagcgtgtccatccatcctcctgtgtgtcgtcgtggt 840
DB	781	GGAACGTCAACCATGCTGTGCAAGCTGTTCATCCATCTCTTGTGTGCTCATCGTGGGT 840
QY	841	tccttggtgtgaagcacttcctccaaaggtcaagcgtgtgtgaactatctggagagac 900
DB	841	TCCTTGTTGTGAAGCACTTCTCCAAAGCTTAACGCGCTGCTGAACATATCTGGACCGAG 900
QY	901	ctataactggagagacacttaactcttattgtctgctataactgttaactgtttaaga 960
DB	901	CTATAACTGGAGAGACCTTAACCTTTATGCTGCTTATACTGTTAACGTGTTTAAAG 960
QY	961	tgaacatacaactatatttctcaagccttcccaagttcaaaccaactataaagcttca 1020
DB	961	TGAACATCAACATATTTTCTACAGCTTTTCCAACTTCAACCAATCTATAAGCTTCA 1020
QY	1021	ttgtcttcttggaatgacgtcgtcttggtcgggtttagatataactaaacacttaacc 1080
DB	1021	TTTGTCTTTGGAAATGATCTGCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1080
QY	1081	tctgcacgttgatcatgtgttcaatataacttcaacttcaacttcaacttcaacttca 1140
DB	1081	TCGCGCATGTGCTCATGTGTCAGTCAATATACTTTCACAACTTAAACAAATATTTAT 1140
QY	1141	ttataatgatatatactcgtacataaagattgttttccagcagctgttccaag 1200
DB	1141	TTTATATATGATATATTTCTGATCAATTAAGATGTTTTTTCGCCAGCGTGTTCATAG 1200
QY	1201	tactagataactaaactgttataactatcttccatgatgatgaacgtccttaagtttta 1260
DB	1201	TACTAGATATTAACCTGTTATTAACCTTATTTCCATGATGTCATGCTTAAAGTTTAA 1260
QY	1261	tgcagaataaattatataatta 1283
DB	1261	TGCAGAAATAAATTATATATTA 1283
RESULT 2		
LOCUS		GM0275972 1355 bp mRNA linear VRT 04-APR-2000
DEFINITION		Gadus morhua mRNA for uracil-DNA glycosylase precursor (ung2 gene).
ACCESSION		AJ275972
VERSION		AJ275972.1 GI:7413913
KEYWORDS		precursor; ung2 gene; uracil-DNA glycosylase.
SOURCE		Atlantic cod.
ORGANISM		Gadus morhua
REFERENCE		1 (bases 1 to 1355)
AUTHORS		Lanes, O. and Willassen, N.P.
TITLE		Identification, cloning and expression of nuclear and mitochondrial uracil-DNA glycosylase from Atlantic cod(Gadus morhua).
JOURNAL		Characterisation of the cold-active catalytic domain
AUTHORS		Unpublished
REFERENCE		2 (bases 1 to 1355)
JOURNAL		Lanes, O.
TITLE		Direct Submission
JOURNAL		Submitted (02-MAR-2000) Lanes O., Department of Biotechnology, University of Tromsøe, Institute of Medical Biology, Medical Faculty, N-9037, NORWAY
FEATURES		Location/Qualifiers
source		1..1355

```

/organism="Gadus morhua"
/db_xref="taxon:8049"
/tissue_type="liver"
/dev_stage="adult"
/country="Norway"
/gene="ung2"
/cds
  /gene="ung2"
  /gene="ung2"
  /EC_number="3.2.2.3"
  /codon_start=1
  /product="uracil-DNA glycosylase"
  /protein_id="CAB85708.1"
  /db_xref="GI:7413914"
  /translation="MIGQOHINSEFSPVSKRKRVSKELEKTEKHAEEVOITPKLRSSN
VEOKTSPOLSEVDELMKAKKALDKIRKATPAGFETWRELAEPKPYKOL
MFADESRHRYVPADQVYSPTEKCDIOGVKYVILIDDPYHGNQAHGCFSTOKP
VPPPSLVNIKELCTDIDGRHGHSLGSMAROGVILLNAVILTRAHQNSHSDRG
METPTDVIKLSVNRSEVFLMGSYAHKKGATIDRRHHVLAQVHPSPLSAHNGFL
GCKHPSKANGLKLSGTEPIPMRAL"
BASE COUNT      393 a      287 c      304 g      371 t
ORIGIN
Query Match      90.7%: Score 1164.2; DB: 5; Length 1355;
Best Local Similarity 98.5%: Pred. No. 1.7e-288;
Matches 1175; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
  QY  91 aaacttatgtttcttaattatgaataacgcggaagaactgaagtcctcaaatg 150
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  DB  163 AACCGAAGAGATGCGGAGAGAGATTGACATATACCGGAGAAAGAGAGTCTCTCAATG 222
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  QY  151 tggaaacaaagacgtcatcgcacacagcttcagtggaagcagctggaaagaatggccaaa 210
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  DB  223 TGGACCAAAAGAGCTATCGCCACAGCTTTCAGTGAGCAGCTGGAAGAATGGCCAAA 282
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  QY  211 ataaagaagcagcgtttgaagaattagcagaagaacagcgtgaggttcogagaaga 270
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  DB  288 ATTAAGAAAGACAGCGCTGACAAAGATTAGACAAAGCAAGCGCTTGAGGTTCGGAGAGA 342
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  QY  271 ctggagaagaagcgtgctgcagagcttgaaaagccacttcaacaacttgatgctc 330
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  DB  343 CTTGGAGAAAGAGCTGCGTGCAGAGTTTGAAAGGCCATCTTCAACATATTGATGCTCT 402
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  QY  331 ttgtgctgaatgaagagcgctcacacgcgtctaccacggcgtgataagtgtagt 390
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  DB  403 TTGTAGCTGATGAGAGAGACCGCTCACACCGCTTACCACCGCTGATCAAGTGTACAGTT 462
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  QY  391 cgacagaagtggtgacattcaagaatgtgaagaagtgtattcttaagccagagcccttacc 450
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  DB  463 GGACAGAGATGTGTACATTCAAGATGTAAAGTGTATTCTTAGGCCAGGACCCCTTACC 522
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  QY  451 acggtcccaacacagacatgtagctgtgttcagtgtaaaagcaggttcccccctccc 510
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  DB  523 ACGGTCCCAACCAAGACATGAGACTCTGTTTCAGTGTGAAAAGCCAGTTCCCCCTCCCC 582
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  QY  511 ccagtcctcgtgaacatatcaagaatgtgtacagacattgtagtgcctcaagcatcctg 570
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  DB  583 CCAGTCTCTGTGAACATATTCAAAGAATGTGTACGACATTTGATGGCTTCAAGCATCTCG 642
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  QY  571 gacatggagaattgaagcgaattgcaaaaacaaagggtgtgtctgtcttaacgggtgtgta 630
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  DB  643 GACATGGAGATCTAAGCGGATGGCAAAACAGAGGCTGCTGCTTAACCGCGGTCTTA 702
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  QY  631 ccgtgctgggcccatacagcacaactccacaagaagcagctgtagagaccttcaaccgacg 690
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  DB  703 CCGTGGGGCCCATCAGGCGCAACTCCACAAAGACAGAGAGCTGGGAGACCTTCAOCGAGG 762
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  QY  691 ctgtgatacaatgtgctgagcgtcaacccgggaagagtgcttctcctgtgtgtgggctcat 750
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  DB  763 CTGTGATCAAGTGCCTGAGCCGTCAACCGGGAAGGAGTGCTTTCCCTGCTGGGCTCAT 822
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  QY  751 acgccataaagaaggagagaccatcgacaggaagaacgltacacatgtcttgcaagctgttc 810

```

```

|||||
DB  823 ACGCCATTAAGAGAGGAGGACCATTCAGAGAGAAAGTGCACATGCTTGAAGCTTTC 882
|||||
QY  811 atccatccctctgtctgctcatcgtgggtctccttggttgtaagaccttccaaagcta 870
|||||
DB  883 ATCATCTCTCTTGTGTGCTCATCGGGGTCTCTTGTTGTGAAGCATCTTCCAAAGCTTA 942
|||||
QY  871 acgggctgtgaacatcttggagcggacataaacttgagaagcactcaactctta 930
|||||
DB  943 ACGGCGTGTGAACATATCTGGGACGAGCCATTAACATGAGAGACATCTAATCTTTTA 1002
|||||
QY  931 tctgtctatactgttactgttlttaagaatcaactcaactatatttctacagctt 990
|||||
DB  1003 TCGTGCCTTAATCTGTAACTGTTTAAAGATGACATCACTATATTTTCTACACCTTT 1062
|||||
QY  991 tccaactcaaacccaatcatalaagcttcatctgtcttcttggaatgctgcttgg 1050
|||||
DB  1063 TCCAACTTCAACCATCTATATAGCTTTTATTTGCTTTTGGAAATGATGCTGTTTGG 1122
|||||
QY  1051 cgggtttagatacttaaacacttaccacactgcacagtgltgacatgltcaagtaata 1110
|||||
DB  1123 CGGTTTAAATCTTAAACACTTTTACCATCTGCGCATGTGACATGTTGAGTCAATA 1182
|||||
QY  1111 taacttcaacttgaacaaaatgttattatattgattatattcttgatcatataa 1170
|||||
DB  1183 TAACTTCAACACTTGAACAAAAGTTATTTATATGATTATATTCGTACATTAA 1242
|||||
QY  1171 gatgttttttccacagcgtgttctcaagtgactaagatactaaactgttataactat 1230
|||||
DB  1243 GATGTGTTTTTCCCGAGCGCTTTTCATAGTAGTACTAGATATTAATGTTATTAACCTAT 1302
|||||
QY  1231 ttccatgatgtcaactgcttaagtttctatgcagaataaatatataattta 1283
|||||
DB  1303 TTTCCATGATGTCAACTGCTTAAGTTTATGACAGAAATTAATATATATTTA 1355
|||||

RESULT 3
MMU55041
LOCUS      MMU55041      1857 bp      mRNA      linear      ROD 25-MAY-1997
DEFINITION Mus musculus uracil-DNA glycosylase (ung) mRNA, complete cds.
ACCESSION U55041
VERSION   U55041.1 GI:1762317
KEYWORDS
SOURCE     house mouse.
            Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE  1 (bases 1 to 1857)
            Svendsen, P.C., Yee, H.A., Winkfein, R.J. and van de Sande, J.H.
            The mouse uracil-DNA glycosylase gene: isolation of cDNA and
            genomic clones and mapping ung to mouse chromosome 5
            Gene 189 (2), 175-181 (1997)
JOURNAL   97311407
MEDLINE   2 (bases 1 to 1857)
REFERENCE 2 (bases 1 to 1857)
            Svendsen, P.C.
            Direct Submission
            Submitted (15-Apr-1996) to Pta C. Svendsen, Medical Biochemistry,
            University of Calgary, Rm.2153 HSC, 3330 Hospital Dr. N.W.,
            Calgary, AB T2N 4N1, Canada
FEATURES
            source
            1..1857
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /chromosome="5"
            1..27
            /gene="ung"
            1..1857
            /gene="ung"
            28..915
            /gene="ung"
            /function="DNA repair enzyme"
            /codon_start=1
            /product="uracil-DNA glycosylase"

```

BASE COUNT	371 a	550 g	441 t
ORIGIN	495 c		
polyA_signal	/gene="ung" 1836 .1841 /gene="ung"		
3' UTR	916 .1857 KANELORSGKKPINMKEL"		

Query Match	28.4%	Score 364.2	DB 10	Length 1857
Best Local Similarity	66.8%	Pred. No. 1e-82		
Matches 519; Conservative	0	Mismatches 258;	Indels 0	Gaps 0;

RESULT	4
LOCUS	BC011039
DEFINITION	BC011039 2002 bp mRNA linear ROD 30-JUL-2001
ACCESSION	Mus musculus, uracil-DNA glycosylase, clone MGC:13929
VERSION	IMAGE:4009947, mRNA, complete cds.
KEYWORDS	BC011039
SOURCE	BC011039.1 GI:15029659
ORGANISM	MGC. house mouse. Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthelia; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 2002)
AUTHORS	Strausberg, R.
TITLE	Direct Submission
JOURNAL	Submitted (25-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC) Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK	COMMENT
NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>	
Contact: MGC help desk	

cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbiology.org>

Clone distribution: MCC clone distribution information can be found through the I.M.A.G.E Consortium/INL at: <http://image.inl.gov>  
Series: IRAK Plate: 18 Row: e Column: 22  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA g1: 18508075.

```
MMTV-LTR/mtl model. Expression driven by an MMTV-LTR
enhancer."
/cione_lib="NCI_CGAP_Lu30"
/lab_host="DH10B"
/note="Vector: pCMV-Sport6"
126..1046
CDS
```

BASE COUNT ORIGIN	417 a	549 c	581 g	455 t
----------------------	-------	-------	-------	-------

Query Match	28.34	Score 302.6	DB 10	Length 2002
Best Local Similarity	66.78	Pred. No. 2.7e-82		
Matches 518; Conservative	0	Mismatches 259	Indels 0	Gaps 0

150 gtggaacaaagacgtcatcgccacagcttctcagtgagcagctggaagaatggtccaa 209

D0 2/3 GUGAGCAGAAACGAGCCAAAGGCATCGCCGCTCAAGCUCUCCGAGAGCAGCUCGACUCCGACUCCAGAGUU 332

QY 210 aataaagaaagcagcgcttgcacaagatttagagcaaaagcaacgacctgaagtttcgagag 269



[illegible][illegible]

Db	804	CACCCCTCCCGCTTCGGTGTACAGAGAGGTTCCCTGGGATGCAGACATTTTCT7AAAGCC	863
QY	870	aacggcgtcgtgaacctactctggacgcggagccctataaactggagagcactctaactc	926
Db	864	AATGAACGTCTCCAGAGACTGTGGCAAGAAGCCCATCACTAGGAGAGACTGTGACCC	920
RESULT	6		
LOCUS	MM08975	1951 bp	linear
DEFINITION	M.musculus mRNA for uracil-DNA glycosylase.		
ACCESSION	Y08975		
KEYWORDS	Y08975.1 GI:1850875		
SOURCE	ung2 gene; uracil-DNA glycosylase.		
ORGANISM	house mouse.		
REFERENCE	Mus musculus		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
JOURNAL	1 (bases 1 to 1951)		
TITLE	Nilsen, H.		
REFERENCE	Submitted (21-OCT-1996) H. Nilsen, UNIGEN Center for Molecular Biology, University of Trondheim, N-7005 Trondheim, NORWAY		
AUTHORS	2 (bases 1 to 1951)		
TITLE	Nilsen, H., Otterlei, M., Haug, T., Solum, K., Nagelhus, T.A., Skjorpen, F., and Krokan, H.E.		
TITLE	Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from different positions in the ung gene		
JOURNAL	Nucleic Acids Res. 25 (4), 750-755 (1997)		
FEATURES	97169285		
source	Location/Qualifiers		
gene	1..1951		
CDs	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/clone_11b="Unizap-XR"		
	69..989		
	/gene="ung2"		
	69..989		
	/gene="ung2"		
	/codon_start=1		
	/product="uracil-DNA glycosylase"		
	/protein_id="CAA70168.1"		
	/db_xref="GI:1850876"		
	/db_xref="WGI:109352"		
	/db_xref="SPTREMBL:P97285"		
	/translation="MICOQTLTSEFSPPTGKRTTRSBEPYSGVALETGGDAVASP AKKRAVNEEGSGPLSAEOLYRIORNKAAALRLRAARVNPFGESWKOOLCGEGR FYFLVMGVFAEENHHKVVYPPPEGVFMTCMDLRDKVYVTLGGDPYHGMQAGLGE SVQDPVPRPSTLENIPEKELSTDDIGFVFLMGSTYIAOKKSDVIRKHHVLIQTAAHPSLVS HKRGMEQFTDAVYSMLNQNLSGVLFLMGSTYIAOKKSDVIRKHHVLIQTAAHPSLVS YRGLGCKHEFAKNEELLQKSGKRINMKEL"		
BASE COUNT	416 a 523 c 565 g 447 t		
ORIGIN			
Query Match	28.1%; Score 361; DB 10; Length 1951;		
Best Local Similarity	66.5%; Pred. No. 7e-62;		
Matches 517; Conservative	0; Mismatches 260; Indels 0; Gaps 0;		
QY	150	gtggacaaaagaacgctcgcgcacagcctttcaagcggagcagctgaaagaatggccaa	209
Db	216	GTCGAGCAGCAAGACGACCAAGCGCGCCGTCAAGCGCCGACGACGACGACGACGACG	275
QY	210	ataagaagacgagccttgacaagaattgagacaaagacgaagcagctgcaggttcggag	269
Db	276	AACAAGGCGCGCGCTGCTCAGGCTCCCGCCGCCAACGTCGCGCGGCTTCGGCGAG	335
QY	270	acttgagagaaagagcctgctgcagagtttgaagaagcactttaaacaatgtatgtcc	329
Db	336	AGCTGGAAGCAGCAGCTGTGCGGGAGATTTCGGGAAGCCGTACTTGTCAAGCTTAATGG	395
QY	330	tttgagctgtatgaaggagccgttaacacgcgtctacccacgcgcgtcatcaatgtacagt	389

Db	Accession	Version	Keywords	Source	Organism	Reference Authors Title Journal	Remark Comment
Db	336	TTGTGCTGGAAGAAAGAACACCACCAAGGCTCTATCCGCCCCCGGACAGGTGTTTCA	455				
Qy	330	tcgaacagagatgctgagacatcaagaatgltgaagatgatcttagagccagaccccttac	449				
Db	456	TGGACCCAGATGTGGCAGATCCGACATGGAAGGTGTTGATCTCTGGACAGATCCCTAT	515				
Qy	450	caaggtcccaacaaacagacatgagctctgtttcagtgfgcaaaaacagctccctccc	509				
Db	516	CACGAGCTTAATCAACAGCTCACGGGCTGTGCTTCAGTGTCCAAAGACCAAGTCCACCCCG	575				
Qy	510	ccagctctcgtbaacatalatacaagaatgltgttacggcatgtatggttcaagcatcct	569				
Db	576	CCCACTTTTGGAATAATTTTAAACAGCTGTCTACAGACATGATGGTTTGTCATCTCT	635				
Qy	570	ggacatgtagatctaaagcggatgggcaaaacaaagggtgtctgtctaaagcgggtctg	629				
Db	636	GGCCCTATGGGGAATTTTTCAGGGGTGGGCCGACACAAAGGTGTCTCTCTCAACGCCCTCTC	695				
Qy	630	accgctggggcccatcaagcgaacaccccaacaaagcaagagcttggagaccttcacagac	689				
Db	696	ACTGTCCCGCCGCCACCAAGCCATTTCCATAAGAGAGAGGGCTGGAGACAGTTCAGGAC	755				
Qy	690	gctgtgatacaagctgagctgagcgtcaaccgggaaggagtctgttcctgtltgagggtca	749				
Db	756	CGGTCGTGTCTCGCTGCTGATCAGAACCTGATGTGGCTCTGTCTCTCTCTGGGGGCTCT	815				
Qy	750	tacggccataagaaggagcgacacctgcacaggaagacgtccacatgttttgaagctt	809				
Db	816	TATGCTCAGAAAGGAGGAGTCTATGTGACAGGAAGCGTCACACATGTTCTGCAGACAGCT	875				
Qy	810	catcatctctcttctgcatcatcgttgggttcttctgttltgtlaagcaacttctcaagagct	869				
Db	876	CACCCCTCCCGCTCTGCTGATGACAGAGGTTCTCTGGATGCAGACATTTTCTTAAAGCC	935				
Qy	870	aacgagctgctgaaactatcttggagcggagacgtataaacttgaagagcaactcaactc	926				
Db	936	AATGAACTGCTCCAAAGTCTGGCAGAGAGCCCACTCACTGGAAGAGACTGTGACCC	992				
RESULT	7						
LOCUS	BC015205	2045 bp	mRNA	linear	PRI 29-OCT-2001		
DEFINITION	Homo sapiens, uracil-DNA glycosylase, clone MGC:17032						
ACCESSION	BC015205						
VERSION	BC015205.1						
KEYWORDS	MGC.						
SOURCE	human.						
ORGANISM	Homo sapiens						
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.						
TITLE	1 (bases 1 to 2045)						
JOURNAL	Strausberg, R.						
	Direct Submission						
	Submitted (01-OCT-2001) National Institutes of Health, Mammalian						
	Gene Collection (MGC), Cancer Genomics Office, National Cancer						
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,						
	USA						
REMARK	NIH-MGC Project URL: <a href="http://mgc.ncl.nih.gov">http://mgc.ncl.nih.gov</a>						
COMMENT	Contact: MGC help desk						
	Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a>						
	Tissue Procurement: ATCC						
	cDNA Library Preparation: Life Technologies, Inc.						
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILMN)						
	DNA Sequencing by: Sequencing Group at the Stanford Human						
	Center, Stanford University School of Medicine, Stanford, CA 94305						
	Web site: <a href="http://www.shgc.stanford.edu">http://www.shgc.stanford.edu</a>						
	Contact: (Dickson, Mark) <a href="mailto:mdickpaxil.stanford.edu">mdickpaxil.stanford.edu</a>						
	Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,						
	R. M.						

Clone distribution: MGC clone distribution information can be found



Oy	459	aaccacagacatggagctctgtttcaagtgtgcaaaagccagttccccctccccccagttcc	518
Db	548	AATCAAGCTCAGGGGCTCTGCTTTAGTGTTCAAAGGCCGTTCCGCGCTCCGCCAGTTTG	607
Oy	519	gtgaacatacataaagaattgttctacgcagcatctgttgcttcaagaatcccttgaaatgta	578
Db	608	GAGAACATTTTAAAGAGTGTCTACAGACATAGAGGATTTTGTTCAATCCTCGGCATGGA	667
Oy	579	gattcaaacggtatgtggcaaaaagaagggtgtcgtcgtcttaacgcggtgtgtacggtcg	638
Db	668	GATTTATCTGGGTGGGGCCAGACGAAAGTGTCTCCCTTCCAAAGCGTGTCTCCAGGTTCT	727
Oy	639	gcccatcagagccaactcccaagaagacagaaggtctggagacatcttaaccagacgttgatc	698
Db	728	GCCCATCAAGCCAACTCTCATAGGAGGAGGAGGCTGGAGCACTTCACTGATGACAGTTGG	787
Oy	699	aagtgtcgtgaagcgttaaccgggaagaagtagtcttctccctgtgttgaggtcatalcagccat	758
Db	788	TCTCTGGCTTAATCAACGACTCGAATGCGCTGTGTTCTTGCTGGGGCTCTTATGCTCAG	847
Oy	759	aagaagggagcgagcatctgcagaggaagaagctacatctctgtcgaagctggtatcatcat	818
Db	848	AAGAAGGGCAGTGGCCATTGATAGGAAGCGGCGACCACTGTACTACAGACGGGCTATGCCCTCC	907
Oy	819	cccttgcctgtcatalcgttgagttctctgtgtgttgaagcaatcttccaaggctaaagctg	878
Db	908	CCTTTGTCAGTGTATAGAGGGTTCTTTGGATGTAGCACTTTTCAAGACCAATGAGCTG	967
Oy	879	ctgaaactatctgtagagagagcctataaacttgagaaagacactctaa	923
Db	968	CTGCGAAGATCTGGCCAAAGACCCATTGCACTGGAAGGAGGACTGTGA	1012

	RESULT	PRI
LOCUS	HS090008	2057 bp linear PRI 15-Apr-1997
DEFINITION	H.sapiens mRNA for uracil-DNA glycosylase.	
ACCESSION	V09008	
VERSION	V09008.1 GI:1850820	
KEYWORDS	ung2 gene; uracil-DNA glycosylase.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 2057)	
AUTHORS	Nilsen, H., Otterlei, M., Haug, T., Solum, K., Nagelhus, T.A., Skorpen, F., and Krokan, H.E. Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from different positions in the UNG gene	
TITLE	Nucleic Acids Res. 25 (4), 750-755 (1997)	
JOURNAL	Medline 97169285	
MEDLINE	2 (bases 1 to 2057)	
REFERENCE	Nilsen, H.	
AUTHORS	Direct Submission	
TITLE	Submitted (22-OCT-1996) H. Nilsen, UNIGEN Center for Molecular Biology, University of Trondheim, N-7005 Trondheim, NORWAY revised by author	
JOURNAL	location/Qualifiers	
REMARK	1..2057	
FEATURES	/organism="Homo sapiens" /db_xref="taxon:9606" /clone_id="Unizap-XR" 71..1012 /gene="ung2" 71..1012 /gene="ung2" /codon_start=1 /product="uracil-DNA glycosylase" /protein_id="CAA70211.1" /db_xref="GI:1850821" /db_xref="SPTRMBL:Q93028" /translation="MGOKLYSFSSPARKRHAPSBEPAVCGTGVAGYEEESGDA"	
CDS		
gene		

AIPKKKAPAGEEPTPEPSSLSLAEODRIORNKAAALLRLRAANVNVGGESEKRLIL  
 SGCEKREYITKLMEVAEERKHVYVPRPHQVFMQMCIDIKVYVILGODPYHGPN  
 OAHGLCEYQVRPPRPSPLENLYKELSTYDIEDFPIVPHGSHLGSAGKQGVLLNANLVY  
 RABDANSKKEKGEQOLFDTAVYSLNQNSSNGVLELWYSYAKQKGSALDRKHNVLQTA  
 HPSPLSVYRGFFGGRHFSKTELLLOKSGKRPTEL

488 a 489 c 539 g 541 t

BASE COUNT  
 ORIGIN

Query Match	27.7%	Score 355.4	DB 9	Length 2057
Best Local Similarity	66.5%	Pred. No. 2e-80		
Matches 509	Conservative 0	Mismatches 236	Indels 0	Gaps 0
Qy 159	aagaatcatcgcccaagcatttcaatgagcaagctggaagaatgcccataaaga	218		
Db 248	ACGCCGCCCTCTCCGCCCTAGTCCACAGCAGTTGGACCGATTCCAAAGAACAGGCC	307		
Qy 219	gcaagcatttgaanaagattagaacaaagcaacgctcgagtltcggaagactgga	278		
Db 308	GGCGCCCTGCTCAGACTCGCGGCCGCAACGCGCCCGGCGCTTGGAGAGAGCTGAAG	367		
Qy 279	agaagctcgctgcagagatttgaanaagccatacttcaacaatgatgtcctttgact	338		
Db 368	AAGCCCTCAGCGGGGAGTTTGGGAAACCGATTATTATCAAGCTAATGGATTGGTGCA	427		
Qy 339	gatgagaagagocgctcacacgcgttaccaccacgcgctgaatgaagttaagctgcagag	398		
Db 428	GAGAAAGAAAGCATTACACTGTTTATTCACACCACCACACCAAGTCTTCACTCGAACCAAG	487		
Qy 399	atgtgtacattcaagaatgtgaagtagtatttctaagcccaagacccttcaacaggtccc	458		
Db 488	ATGTGTGACATATAAAGATGTGAAGTGTGTCAATCCGTGGAGACGAGATTCATATGAGACT	547		
Qy 459	aaccaagacatgtgactctgtttcagtggttgcataaagccagttcccccctccagctc	518		
Db 548	AATCAGCTCACGGGCTCTGCTTATGTGTTCATAAGGCTGTTCCGCTCCGCCACGTTTG	607		
Qy 519	gtgaacatacaagaagtgtgtacagcaattgtatgcttcaagcatcccggaacatga	578		
Db 608	GAGACATTTTATTAAGAGATTGTTCTACAGACATAGAGATTGTTGTCATCTGGCCATGGA	667		
Qy 579	gattcaagcggatgtggcaaaaacaaaggggtgtctgtcttcaacgcggtgtcagctgcg	638		
Db 668	GATTATCTGGGTGGGCGCAGAGCAAGTGTCTCCCTTCAAGCGTGTCTCAGCGTTGCT	727		
Qy 639	gcccatcagcccaactcccaagaagacagagcttgggaaccttcaaccgacgctgtgac	698		
Db 728	GCCCATCAAGCAACTCTCATTAAGAGAGGAGGCTGGAGCACTGATCAGTACAGTTGTG	787		
Qy 699	aagtgtcgtgagcgttaaacccgggaagagatcgatttccgtgtgtgggtctataagccat	758		
Db 788	TCTCTGGCTTAATTCAGAACCTCAATGCGCTTGTCTTCTGCTCGGCGCTCTTATGCTCAG	847		
Qy 759	aagaagggagagccatctgcacgaggaagctaacctatgtcttgcgaagctgttcatcatct	818		
Db 848	AAGAAAGGCGAGCTCCATTGATGAGAGCGGCACCATGTACTACAGACGGCTCATDCCCTCC	907		
Qy 819	ccttgtctgtcatcagctggtgttcccttggttgtaagcaactctccaaggtcaacggctg	878		
Db 908	CCTTTGTGAGTGTATAGAGGGGTTCTTTTGGATGTACACACTTTTCAAAACCAATGAGCTG	967		
Qy 879	ctgaaactatcttggagcggagctataaacttggagagagcactctaa 923			
Db 968	CTGCAAGAGTCTGGCAAGAGCGCCATTACTGTGAAGAGAGCTGTGA 1012			

RESULT 10

HSUNG

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

Human cDNA for uracil-DNA glycosylase.

X15653.1

GI:37598

DNA repair; UNG gene; uracil-DNA glycosylase.

2062 bp

mRNA

linear

PRI 15-Apr-1997



Db 384 GGGCCCCCTGCTAGACTCCGGCCGCCAAGCTGCCGCTTGAGAGAGCTGGAAAG 443  
 Qy 279 agagagctgctgcagagtttgaagaaagccatacttcaaaacttgatcgttctgtagct 338  
 Db 444 AAGCACTCAGGGGAGGTTGGGAAACGGTATTTATATCAAGCTATAGGATTTGTTGCA 503  
 Qy 339 gatgagagagccgttcaacccgttcaaccagcgctgataagttgacagttcagagag 398  
 Db 504 GAAGAAAGAACCATTTACCTGTTTATCCACCCGCCAACCATCTTCACTTGAGCCAG 563  
 Qy 399 atgtgtgaactcaagaatgtgaagtagtattcttagccagaccccttaacagagctccc 458  
 Db 564 ATGTGTGACATATAAGATGTGAAGGTTGTTCATCTCGGACAGGATCATATTCATGACCT 623  
 Qy 459 aaccaagacacatgtagctcgttctcagtgatgcaaaaagccagcttcccccacagcttc 518  
 Db 624 AATCAAGCTCAGCGGCTCTGCTTACTTACTTCAAAAGCCCTGTTCCGCCCTCCGACGTTTG 663  
 Qy 519 gtgaacataacaaagaattgtgtacacacattgtgcttcaagaacctgagatgga 578  
 Db 684 GAGAACATTTATAAGAGTTGTCTACAGACATAGAGGATTTGTTCATCTCGGACATGGA 743  
 Qy 579 gactcaagggagatggcaaaaagagggatgctgtcgttcaacgggtgctgacagctgagc 638  
 Db 744 GATTATATCTGGGTGGCCAAAGCAAGGTGTCTCTCTCAACGCTTCTCTCAAGGTTTCCG 803  
 Qy 639 gcccatcaagggcaactcccaacagaagagctgaggaagaccttcaacggatgctgac 698  
 Db 804 GCCCATCAAGCAACTCTCATTAAGGAGGAGCGGCTGGAGCAGTTCACTGATGACGTTGTG 863  
 Qy 699 aagtgctgtagcgttcaacccgggaagagtcgttttccgttctgtgtggggtccatagcccat 758  
 Db 864 TCCTGGCTTAATCAGAACCTCGAATGCGCTTGTCTCTCGGCGCTCTTAATGCTCAG 923  
 Qy 759 aagaaggggagccacatcgacagaagacgttcaacacatgcttctgcaagctgttcatcat 818  
 Db 924 AAGAAAGGAGAGTCCATTATATAGGAAGCGGACCATGTAACAGACGGTCTATCCCTCC 963  
 Qy 819 ccttctgtcgtcactcgtgtgttcttctgtgtgtgaagcacttctcaagggctaaaggcgctg 878  
 Db 984 CTTTGTCTAGTGTATAGAGGGTCTTGTGATGTAGACACTTTTCAAAAGCAATGAGCTG 1043  
 Qy 879 ctgaaactatctgggaagcgagctataaactgtgagagcactctaa 923  
 Db 1044 CTGCAAGAGCTGGCAAGAACGCCATTGACTGGAAGAGAGCTGTGA 1088

**RESULT 12**  
 BC004037 1737 bp mRNA linear ROD 12-JUL-2001  
 LOCUS Mus musculus, similar to uracil-DNA glycosylase, clone MGC:7655  
 DEFINITION IMAGE:3495894, mRNA, complete cds.  
 ACCESSION BC004037  
 VERSION BC004037.1 GI:13278470  
 KEYWORDS MGC.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 1737)  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-FEB-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
 COMMENT Contact: MGC help desk  
 Email: [cgabs-r@mail.nih.gov](mailto:cgabs-r@mail.nih.gov)  
 Tissue Procurement: Lohar Hemmighausen Ph.D., Robin Humphreys  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)  
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome

Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www-shgc.stanford.edu>  
 Contact: (Dickson, Mark) [mdickson@stanford.edu](mailto:mdickson@stanford.edu)  
 Dickson, M., Schmitz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>  
 Series: IRAP Plate: 8 Row: 1 Column: 4.  
 Location/Qualifiers  
 1. 1737  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="MGC:7655 IMAGE:3495894"  
 /tissue="Mammary tumor. WAP-TGF alpha model. 7 months  
 old, gross tissue."  
 /clone.lib="NCI CGAP\_Mams"  
 /lab.host="DH10B"  
 /note="Vector: pCMV-Sport6"  
 192..791  
 /codon\_start=1  
 /product="Similar to uracil-DNA glycosylase"  
 /protein\_id="AAH04037.1"  
 /db\_xref="GI:13278471"  
 /translation="MGFAEERNHKKVYPRPQVFTWOMDIRVKKVILIGDPRYHG  
 PMOAGLCSVORVPPPSLENIRKELSTDIDGVHYRCHDLSGMARGVLLNLA  
 TVRAHQANSHKRWGEQFTDAVSWMLNGLSVLFWGSYVQKRGSVYDRKHVQLQ  
 TAPSPPLSYVRGLFCRHSKSNELIOLSKGRKIMKWL"  
 BASE COUNT 397 a 423 c 459 g 458 t  
 ORIGIN  
 Query Match 24.9%; Score 319; DB 10; Length 1737;  
 Best Local Similarity 69.9%; Pred. No. 4, 3e-71;  
 Matches 430; Conservative 0; Mismatches 185; Indels 0; Gaps 0;  
 Qy 312 ttcaacgaattgctgcttctgtgagtgagagagcgcgttaccgcttaccaccg 371  
 Db 180 TTAAACCACTAATGGGATTTGTGCTGAAGAAAGACCCACCAAGCTATTCGCCC 239  
 Qy 372 gctgacaaagtgtacagttcgcagagatgtgtgacatccaagatgtgaagtagtatt 431  
 Db 240 CCGGACAGAGTGTTCACATGAGACCCAGATGTGGGACATCGAGATGTGAAGTGTGAT 299  
 Qy 432 ctaggccaggacccttaccacaggttcccaacaaagacacatgtagctgttcaagtgc 491  
 Db 300 CTGGACAGAGATCTCATATACAGGACCTAATCAAGCTACAGCGGCTGTGCTCACTG 359  
 Qy 492 aagccagttcccccctcccccacatctcgtgaataataaagaatgtgtaccgactt 551  
 Db 360 AGACCAAGTTCACCCCGCCAGATTGGAATAATTTTAAAGCTGTCTACAGACATC 419  
 Qy 552 gatgcttcaagacatctgagatctgaagcggatgagggcaacaagaagggtgctg 611  
 Db 420 GATGTTTGTGTTATCTTGCCATGCGGATTTGTCAAGGATGGGCGGACAGAGTGTCT 479  
 Qy 612 ctgttcaagcgggtgtgacccgtgctgagccatcaagggcaactcccaacagaagagc 671  
 Db 480 CTCCTCAAGCGCGTCTCACTGCTCCGCGCCACCAAGCAATTCATTAAGAGAGGGCG 539  
 Qy 672 tgggaagacttcaacgcagctgtgacaaagtgtgctgaagcgggaaggaagtcgt 731  
 Db 540 TGGGACAGATTTCAGCGAGCGGTCGTGCTGCTGAATTCAGAACCTGAGTGGCTGTG 599  
 Qy 732 ttccctgtgtggggtcattagcccaataagaaggagcgacacatgacaggaagctac 791  
 Db 600 TTCTCTCTCTGGGCTCTTATGCTCAGAAAGAGGCAAGTCTATTATACAGGAAGCTCAC 659  
 Qy 792 catgtcttgaagctgttcatcatctcttctgtcgtcactgtgttctgttctgtt 851  
 Db 660 CAGTGTCTGACAGACACTCACCCCTCCCGCTGTGCTGACAGAGGTTCTCGGATGCG 719  
 Qy 852 aagcacttctccaaggcgaagcgtgtgaactatctggagcggagcctataactg 911

Db 720 AGACATTTTCTAAAGCAGTGAAGTCTCCAGAGTGTGGCAAGAGCCCATCACTG 779  
QY 912 agagcacttaactc 926  
Db 780 AAGGAGCTGTGACCC 794

RESULT 13  
CNS01DWM 189672 bp DNA linear PRI 25-MAY-2000  
LOCUS Human chromosome 5 DNA sequence \*\*\* IN PROGRESS \*\*\* BAC R-391B7 of  
DEFINITION library RPCR-11 from chromosome 5 of Homo sapiens (Human), complete  
sequence.  
ACCESSION AL136059.2 GI:7106618  
VERSION AL136059  
KEYWORDS HTG; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 189672)  
GENOSCOPE.  
TITLE Direct Submission  
JOURNAL Submitted (23-MAY-2000) to the EMBL/GenBank/DBJ databases  
COMMENT On Feb 27, 2000 this sequence version replaced gi:6981976.  
IMPORTANT: This sequence is unfinished and does not necessarily  
represent the correct sequence. Work on the sequence is in progress  
and the release of this data is based on the understanding that the  
sequence may change as work continue. The sequence may be  
contaminated with foreign sequence from E.coli, yeast vector,  
phage, etc. even if efforts are made to eliminate these  
contaminating sequences. The following BAC sequence is oriented  
from the 77 to the 526 end.

-----  
Overall quality chart :  
Range : bases  
0 : 2  
1 - 9 : 231  
10 - 19 : 880  
20 - 29 : 2766  
30 - 39 : 7567  
40 - 49 : 8585  
50 - 59 : 13484  
60 - 69 : 28999  
70 - 79 : 55767  
80 - 89 : 42661  
90 - 99 : 28730

Percentage of bases with a quality value >= 40 : 93 %.

FEATURES  
source  
1.189672  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone.lib="R-391B7"  
/clone.R="R-391B7"  
BASE COUNT 57335 a 40587 c 38988 g 52760 t 2 others  
ORIGIN

Query Match 20.5%; Score 262.8; DB 9; Length 189672;  
Best Local Similarity 62.8%; Pred. No. 4.3e-56;  
Matches 491; Conservative 0; Mismatches 282; Indels 9; Gaps 5;

QY 169 cgcacagcttccagtcgagcagtcgaaagtcgcaaaataagaagcagcgttg 228  
Db 24120 CTCCTCAGACAGAGAGAGAGAGTATCTGTCTCCAGAGAAACATGCGCACAGCCCTGC 24179  
QY 229 acaagattgagcaaaagcagcgtcgtcaggttcgagagagacttgagagagagcgttg 288  
Db 24180 TCAGACTCACACCTACACATGCTC-GTGGGCTTGTGTGAGATTGTGAGAGAGCATCTCA 24238

QY 289 ctgcagagttgaaagccacttaacaaatgaltgtccttgtgagtcgtagagga 348  
Db 24239 GCAGAGATTTGGGAAACCATATTTTAAAGCTCATGATGTGTGACGATGAAAGCA 24298  
QY 349 gccgtcacaccgtctaccacccggtcgtatcaagtgatcagttgacagagatgtaga 408  
Db 24299 AGGATTACACTGTTATCTCTCCACACCAAAATATTCACCTGAGCTAAGTGTGACA 24358  
QY 409 ttcagatgtgaaagtagtgcattagcgccagaccccttaacagc--tcccaacagc 466  
Db 24359 TAAGAGATGTGAAGAGATGTCATTCAGAGAGAGATTCATATGAGACCCCATCAAGC 24418  
QY 467 acatgaccttgcttccagtgctgcaaaagccaggttccctcccccagctcgtgtaacat 526  
Db 24419 TCGTGGGTCTGTTAGTGTCAAAAGCCTGTTCACCTGCTGCCCGCTTGGAAAAAT 24478  
QY 527 atcaagaatgtgtacacattgagcttcaagccttcagcctcggaatgagatcgaag 586  
Db 24479 TTATTAAGAGACTGTCTACACATAGATGCTTTGTTCAATCCAGCCATGAGATGTATC 24538  
QY 587 cggatggcacaacaaaggggtgctgcttaacgcggtgctgacgctgcggccatca 646  
Db 24539 TGGGTGGGTCAAAACAAAGTG--TTTACTCAACGCTGTCTCAGTCCAGCCATCA 24595  
QY 647 ggcacactcccaagagcagcagcgtgag--acctcacgacgtctgacgaagtg 704  
Db 24596 AGCCAAATTCATGAG 24655  
QY 705 ctgagcgtcaacgggagagagagctgttctcgttctggtggggtcattacgccaataag 764  
Db 24656 CTAAAGCCAGAACCTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 24715  
QY 765 gtagcagcactcagcaggaagcgtcaccatgcttcaagcgttcaatcactcctttg 824  
Db 24716 ACAAGT--CCATGAG 24774  
QY 825 tctgctatcgtgtgtccttctgttgaagcacttcccaagcgttaacggtcgtgtaaa 884  
Db 24775 TCAGCGTATGATGAGTGTCTTGTGATGAGACCTTTCTTAAGACCAATGAGAGAGAGAG 24834  
QY 885 ctatctgagcagcagcctataaactgagagacacttaactcttaatgctccttact 944  
Db 24835 AAGTCAGCAGCAAG 24894  
QY 945 gtc 946  
Db 24895 GT 24896

RESULT 14  
AC010258 213353 bp DNA linear HTG 20-APR-2001  
LOCUS Homo sapiens chromosome 5 clone CTC-452117, WORKING DRAFT SEQUENCE,  
DEFINITION 17 unordered pieces.  
ACCESSION AC010258  
VERSION AC010258.3 GI:13699482  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 213353)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 5  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 213353)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT On Apr 20, 2001 this sequence version replaced gi:9964744.  
-----Genome Center  
Center: Joint Genome Institute







---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 22, 2002, 03:54:35 ; Search time 270.17 Seconds

(without alignments)  
8153.389 Million cell updates/sec

Title: US-09-758-017A-1

Perfect score: 1283

Sequence: 1 gacatccgctgcaaatatg.....agaataaatatatttta 1283

Scoring table: IDENTITY\_NUC

Searched: 1736436 segs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : \_N\_Geneseq\_032802:\*

1: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1278.2	99.6	1283	22 AAS09498
2	1164.2	90.7	1355	22 AAS09499
3	335.4	27.7	1967	22 AAH34344
4	335.4	27.7	2053	18 AAT73564
5	335.4	27.7	2122	24 AAS94978
6	207.6	16.2	1532	22 AAS63241
7	180	14.0	657	22 AAF94351
8	180	14.0	1830121	17 AAT42063
9	172.2	13.4	349980	21 AAF21609

10	172.2	13.4	349980	21 AAF21610
11	160.8	12.5	2469	20 AAX13540
12	159.6	12.4	681	20 AAX27774
13	147.8	11.5	8753	12 AAH3285
14	142.2	11.1	1348	21 AAG36561
15	138	10.8	2944528	24 ABA03041
16	137.4	10.7	11340	13 AAV52280
17	132.2	10.3	2988	23 AAS75904
18	121.4	9.5	17846	18 AAV74420
19	111.4	8.7	2975	22 AAH53981
20	111.4	8.7	4114	22 AAH54820
21	111.2	8.7	651	22 AAH52516
22	107.4	8.4	133719	21 AAC64754
23	107	8.3	335100	20 AAV78803
24	107	8.3	137507	19 AAV19941
25	106.4	8.3	487	21 AAC78326
26	102	8.0	4100	20 AAZ32019
27	102	8.0	4100	22 AAC90076
28	100	7.8	119211	22 AAF28553
29	97.8	7.6	910715	20 AAX20248
30	94	7.3	1230025	20 AAX91990
31	89	6.9	4403765	22 AAI99683
32	89	6.9	4411529	22 AAI99682
33	83	6.5	7361	19 AAV62153
34	83	6.5	8852	19 AAV62130
35	83	6.5	117213	19 AAV62176
36	83	6.5	154746	24 AAD25519
37	80	6.2	500	22 AAH41150
38	78.2	6.1	4881	19 AAV63438
39	68	5.3	124884	22 AAH74201
40	68	5.3	125157	22 AAH74202
41	67.4	5.3	777	19 AAX14447
42	61.6	4.8	13228	23 AAS59567
43	49.4	3.9	762	22 AAH66426
44	49.4	3.9	349980	22 AAH68528
45	48.4	3.8	1262	23 AAS65031

#### ALIGNMENTS

RESULT 1	
AAS09498	
ID AAS09498 standard; CDNA: 1283 BP.	
XX	
AC AAS09498;	
XX	
DT 24-OCT-2001 (first entry)	
XX	
DE Atlantic cod cDNA encoding heat-labile uracil-DNA glycosylase, UNG #1.	
XX	
KW Atlantic cod; heat-labile uracil-DNA glycosylase; UNG; UDG;	
KW PCR control; LCR control; ligase chain reaction; carry-over prevention;	
KW SS.	
XX	
OS Gadus morhua.	
XX	
XX	
FT Key	Location/Qualifiers
FT CDS	18..923
FT	/*tag= a
FT	/Product= "UNG #1"
XX	
PN WO200151623-A1.	
XX	
PD 19-JUL-2001.	
XX	
PF 10-JAN-2001; 2001WO-NO00008.	
XX	
PR 12-JAN-2000; 2000NO-0000163.	
XX	
PR 27-OCT-2000; 2000NO-0005428.	
XX	
PA (BIOT-) BIOTEC ASA.	
XX	

Nelisseria meningit  
Enterococcus faec  
B. pallidus uracil  
P. denitrificans Ec  
Arabidopsis thalia  
Listeria monocytog  
Streptococcus pneu  
DNA encoding novel  
Staphylococcus aur  
S. epidermidis gen  
S. epidermidis ope  
Macaca mulatta rha  
KSHV LUR DNA (nucl  
KSHV long unique c  
Human cancer assoc  
Human MERT1 relate  
L1385 cDNA clone.  
Genomic fragment #  
Borrelia burgdorfe  
Nucleotide sequenc  
Mycobacterium tube  
Mycobacterium tube  
HSV-2 strain SB5 C  
HSV-2 strain SB5 C  
HSV-2 strain SB5 C  
Human herpesvirus  
Left nucleic acid  
Feline herpesvirus  
Nucleotide sequenc  
H. pylori GHP0 902  
Propionibacterium  
C glutamicum codin  
C glutamicum codin  
DNA encoding novel

PI Lanes C, Willasen NP, Guddal PH, Gjellesvik DR;  
 DR WPI: 2001-451854/48.  
 DR P-FSDB; ANU04939.  
 XX  
 XX  
 PT New cod liver reaction-DNA glycosylase enzyme, useful in monitoring or  
 PT controlling a reaction system multiplying DNA sequences or in  
 PT carry-over prevention procedures -  
 XX  
 XX  
 PS Claim 7; Page 52-54; 59pp; English.  
 XX  
 XX The sequence encodes an Atlantic cod heat-labile uracil-DNA glycosylase  
 CC (UNG/UDG). The enzyme is useful in monitoring and/or controlling a  
 CC reaction system multiplying DNA sequences, e.g. PCR (polymerase chain  
 CC reaction) or LCR (ligase chain reaction). The enzyme is also useful in  
 CC carry-over prevention procedures.  
 XX  
 XX Sequence 1283 BP; 364 A; 278 C; 280 G; 360 T; 1 other;  
 SQ

Query Match	99.68	Score 1278.2	DB 22	Length 1283
Best Local Similarity	99.88	Pred. No. 0		
Matches 1280	Conservative	0	Mismatches 3	Indels 0
				Gaps 0

OY	1	gaatccgcgtgcaaatatggttctgaagttaggttatgccaagatgatatcaatca	60
Db	1	gacacccgcgtgcgaataatgtttgttcaagttaaggtttatgccaagaatgatcatcaaa	60
OY	61	atcgggtgttaccgaagtttacttaattccccaacttatgtttttctcaaatatagaaga	120
Db	61	atcgggtgttaccgaagtttacttaattccccaacttatgtttttttctcaaatatagaaga	120
OY	121	taacgcgcgaagaacatgaggttcctcaaatgttgacaacaagaacgttcatgcgcacagctt	180
Db	121	taacgcgcgaagaacatgaggttcctcaaatgttgacaacaagaacgttcatgcgcacagctt	180
OY	181	cagtgtagcagctgtgaaagaattgccaataaagaagaagcgtttgacaagaatttagag	240
Db	181	cagtgtagcagctgtgaaagaattgccaataaagaagaagcgtttgacaagaatttagag	240
OY	241	caaaagcaaacgcctgcagagttcttcgagaagacttbgagaagaagacgtgcgcagagtttg	300
Db	241	caaaagcaaacgcctgcagagttcttcgagaagacttbgagaagaagacgtgcgcagagtttg	300
OY	301	aaaagccaacttccaacaactatgatgttcctttgactgtatgagaagagacgcgttacaacgcg	360
Db	301	aaaagccaacttccaacaactatgatgttcctttgactgtatgagaagagacgcgttacaacgcg	360
OY	361	tctaccccaacgcgcgtatcaagtgtgtacaggttcgaaagagatggttgacattccaagaatgtga	420
Db	361	tctaccccaacgcgcgtatcaagtgtgtacaggttcgaaagagatggttgacattccaagaatgtga	420
OY	421	aagtagtgatattcttgccgacgaaccccttaccacgcgttcccaaccagaacacatgagctcgtt	480
Db	421	aagtagtgatattcttgccgacgaaccccttaccacgcgttcccaaccagaacacatgagctcgtt	480
OY	481	tcaagtgtgcaaaaagccagtttccccctccccccagttctcgtgtgaacatatatacaagaatgtt	540
Db	481	tcaagtgtgcaaaaagccagtttccccctccccccagttctcgtgtgaacatatatacaagaatgtt	540
OY	541	gtaccgcgaattgtagtgccttcaagaacatcccttgagacatgtgagatctraagcgtatggcaaac	600
Db	541	gtaccgcgaattgtagtgccttcaagaacatcccttgagacatgtgagatctraagcgtatggcaaac	600
OY	601	aaggggtgcctcgtcttcaacgcggtgttcgtgacgcgtgacccatcaaggccaactcccaaa	660
Db	601	aaggggtgcctcgtcttcaacgcggtgttcgtgacgcgtgacccatcaaggccaactcccaaa	660
OY	661	aggaacaagaggtctggagaaccttcaaccgcagcgtgtgatacaagtgtgctlaagcgttcaacccgg	720
Db	661	aggaacaagaggtctggagaaccttcaaccgcagcgtgtgatacaagtgtgctlaagcgttcaacccgg	720
OY	721	aaggaagtgcgtttccctcgtttgttggtgcctatcagcccaatgaagaaggtgacacatcgaca	780

Dh	721	aagagagtggttttcctgtgttggtgggtccataagccataagaaggagggaccatcgcaga	780
Qy	761	ggaaacgctacacatgctcttgcaagctgttcacatccctccttgctgtcactggtggt	840
Dh	781	ggaacgctacacatgctcttgcaagctgttcacatccctccttgctgtcactggtggt	840
Qy	841	tccttggtgtgaagcactctcccaaggtcaagggctgctggaacatctgagcggagc	900
Dh	841	tccttggtgtgtgaagcactctcccaaggtcaagggctgctggaacatctgagcggagc	900
Qy	901	ctataacctggagagcacctcaactctttagctgtcctataactgttatacgtgtttaaga	960
Dh	901	ctataacctggagagcacctcaactctttagctgtcctataactgttatacgtgtttaaga	960
Qy	961	tgaacatcacactatattttctaaagttttccaagtttaaaccaactataagtttca	1020
Dh	961	tgaacatcacactatattttctaaagttttccaagtttaaaccaactataagtttca	1020
Qy	1021	tttgctctttggaatgtagctgcctttgtgcggtttttagataactttaaacactttacac	1080
Dh	1021	tttgctctttggaatgtagctgcctttgtgcggtttttagataactttaaacactttacac	1080
Qy	1081	ctcgacatgctgactcatggttcagtcaataaacttccacactggaacaaaagtat	1140
Dh	1081	ctcgacatgctgactcatggttcagtcaataaacttccacactggaacaaaagtat	1140
Qy	1141	tttataatgtgttatctctgtaantaagaagtgttttttcccgaggtgtttcaatg	1200
Dh	1141	tttataatgtgttatctctgtaantaagaagtgttttttcccgaggtgtttcaatg	1200
Qy	1201	tactagagataataacgtttataaactatttccatagatgtaacatgcttaagtttta	1260
Dh	1201	tactagagataataacgtttataaactatttccatagatgtaacatgcttaagtttta	1260
Qy	1261	tgcagaaataataatatattta 1283	
Dh	1261	tgcagaaataataatatattta 1283	

RESULT	2	
AA509499		
ID	AA509499	standard; cDNA; 1355 BP.
XX		
XX	AA509499;	
AC		
XX		
DT	24-OCT-2001	(first entry)
XX		
DE	Atlantic cod	cDNA encoding heat-labile uracil-DNA glycosylase, UNG #2.
XX		
KW	Atlantic cod; heat-labile	uracil-DNA glycosylase; UNG; UDG;
KM	PCR control; ICR control;	ligase chain reaction; carry-over prevention;
KW	ss.	
XX		
OS	Gadus morhua.	
XX		
EH	Key	Location/Qualifiers
FT	CDS	90..995
FT		/*tag= a
FT		/product= "UNG #2"
XX		
PN	WO200151623-A1.	
XX		
PD	19-JUL-2001.	
XX		
PF	10-JAN-2001; 2001WO-N000008.	
XX		
PR	12-JAN-2000; 2000NO-0000163.	
PR	27-OCT-2000; 2000NO-0005428.	
XX		
PA	(BIOT-) BIOTEC ASA.	
XX		
PI	Lanes O, Willasen NP, Guddal PH, Gjellesvik DR;	

XX WPI: 2001-451854/48.  
DR P-PSDB: AAU04940.  
PT New cod liver uracil-DNA glycosylase enzyme, useful in monitoring or  
PT controlling a reaction system multiplying DNA sequences or in  
PT carry-over prevention procedures  
XX  
XX  
PS Claim 7, Page 54-56; 59pp; English.  
XX  
XX The sequence encodes an Atlantic cod heat-labile uracil-DNA glycosylase,  
CC (UNG/UNG). The enzyme is useful in monitoring and/or controlling a  
CC reaction system multiplying DNA sequences, e.g. PCR (polymerase chain  
CC reaction) or LCR (ligase chain reaction). The enzyme is also useful in  
CC carry-over prevention procedures.  
XX  
SQ Sequence 1355 BP; 393 A; 287 C; 304 G; 371 T; 0 other;

Query Match 90.7%; Score 1164.2; DB 22; Length 1355;  
Best Local Similarity 98.5%; Pred. No. 0;

Matches 1175; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 91 aaacttaagtcttcttaattgaataacgcggaagaactgagctccaatg 150  
DB 163 aaaccgaaaagcatgcccgaagaagttcagataacgcggaagaactgagctccaatg 222  
QY 151 tggaaacaaagacgctcaccagcttcagtgagcagctggaagaatggccaaa 210  
DB 223 tggaaacaaagacgctcaccagcttcagtgagcagctggaagaatggccaaa 282  
QY 211 ataagaagcagcgtctgaagaagtaagcgaagaacagcgtctgagtgaga 270  
DB 283 ataagaagcagcgtctgaagaagtaagcgaagaacagcgtctgagtgaga 342  
QY 271 ctggagaagaagcgtgctgagagttgaaagccatactccaacatgagtcct 330  
DB 343 ctggagaagaagcgtgctgagagttgaaagccatactccaacatgagtcct 402  
QY 331 ttgtatgctgataagagagcggtcacaccgtctaccaccggtctatcaagtgtacgtt 390  
DB 403 ttgtatgctgataagagagcggtcacaccgtctaccaccggtctatcaagtgtacgtt 462  
QY 391 ccacagagatgtgtacatctcaagaatgtgaagtagtgaattgaagccagaccctacc 450  
DB 463 ggaacagagatgtgtacatctcaagaatgtgaagtagtgaattgaagccagaccctacc 522  
QY 451 acggtcccaacccaagacacatgagctgttcaagtgacgaagaacagctcccccctccc 510  
DB 523 acggtcccaacccaagacacatgagctgttcaagtgacgaagaacagctcccccctccc 582  
QY 511 ccaggtctcgttgacataatacaagaatgtgtacacgacatgtatggtctcaagcatctg 570  
DB 583 ccaggtctcgttgacataatacaagaatgtgtacacgacatgtatggtctcaagcatctg 642  
QY 571 gacatggaatcctaaagcgaatgtgcaaaacgaaggtgtgtgtcttaacgsgtgcctga 630  
DB 643 gacatggaatcctaaagcgaatgtgcaaaacgaaggtgtgtgtcttaacgsgtgcctga 702  
QY 631 ccgtgacgagccatcagcgaacaccccaacgaagagcgtggaagacctcaccagcag 690  
DB 703 ccgtgacgagccatcagcgaacaccccaacgaagagcgtggaagacctcaccagcag 762  
QY 691 ctgtgatacagtggtgagcgtcaacccggaagagtgctttctcgtgtgtgggtctat 750  
DB 763 ctgtgatacagtggtgagcgtcaacccggaagagtgctttctcgtgtgtgggtctat 822  
QY 751 agccccaataaggaagagcagccatcagacaggaagaacgtaccatgtctgcaagcgttcc 810  
DB 823 agccccaataaggaagagcagccatcagacaggaagaacgtaccatgtctgcaagcgttcc 882  
QY 811 atccatctccttgtctgctacatcgtaggttctcctgtgtgtgaagcactctccaagccta 870

DB 883 atccatctccttgtctgctacatcgtaggttctcctgtgtgtgaagcactctccaagccta 942  
QY 871 acggtgctgtaaacatacttggagcggagcctataaacctggaagacatccttcta 930  
DB 943 acggtgctgtaaacatacttggagcggagcctataaacctggaagacatccttcta 1002  
QY 931 tctgtcttaactgttgaactgtttaagaatgaacacacatactttcttaacgctt 990  
DB 1003 tctgtcttaactgttgaactgtttaagaatgaacacacatactttcttaacgctt 1062  
QY 991 tccaagtccaacccaatcataaagcttcatgtcttcttggatgatgctgcttgg 1050  
DB 1063 tccaagtccaacccaatcataaagcttcatgtcttcttggatgatgctgcttgg 1122  
QY 1051 cgggtttagatacctaataaacacttaccactctgcacatggttgacatggttcaagtaata 1110  
DB 1123 cgggtttagatacctaataaacacttaccactctgcacatggttgacatggttcaagtaata 1182  
QY 1111 taacttcaacttgaacaaaatgttattttatattgattatattctgtacattaa 1170  
DB 1183 taacttcaacttgaacaaaatgttattttatattgattatattctgtacattaa 1242  
QY 1171 gattgttttttccagcgtgttctcaatgagtaacgataataaactgttataactat 1230  
DB 1243 gattgttttttccagcgtgttctcaatgagtaacgataataaactgttataactat 1302  
QY 1231 ttccatgatgtcaactgcttaagttttatgcagaataaattatattta 1283  
DB 1303 ttccatgatgtcaactgcttaagttttatgcagaataaattatattta 1355  
  
RESULT 3  
AAH34344  
ID AAH34344 standard; CDNA; 1967 BP.  
XX  
XX AAH34344;  
AC  
XX  
XX 03-SEP-2001 (first entry)  
DT  
XX  
XX Human colon cancer antigen encoding CDNA SEQ ID NO:1426.  
DE  
XX  
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;  
KW colorectal carcinoma; chromosome 12; ss.  
KW  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200122920-A2.  
PN  
XX  
XX 05-APR-2001.  
PD  
XX  
XX 28-SEP-2000; 2000MO-US26524.  
PF  
XX  
XX 29-SEP-1999; 990US-0157137.  
PR  
XX  
XX 03-NOV-1999; 990US-0163280.  
PR  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX  
XX Ruben SK, Barash SC, Birse CE, Rosen CA;  
PI  
XX  
XX WPI: 2001-235357/24.  
DR  
XX  
XX P-PSDB: AAG74939.  
DR  
XX  
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
PT useful for preventing, diagnosing and/or treating colorectal cancers -  
PT  
XX  
XX Claim 1; Page 3119; 9803pp; English.  
PS  
XX  
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
CC the proteins are collectively known as colon cancer antigens. The colon  
CC cancer antigens have cytostatic activity and can be used in gene  
CC therapy and vaccine production. N and P may be used in the prevention,  
CC diagnosis and treatment of diseases associated with inappropriate P

CC expression. For example, N and P may be used to treat disorders  
 CC associated with decreased expression by rectifying mutations or deletions  
 CC in a patient's genome that affect the activity of P by expressing  
 CC inactive proteins or to supplement the patient's own production of P.  
 CC Additionally, N may be used to produce the colon cancer-associated Ps,  
 CC by inserting the nucleic acids into a host cell and culturing the cell  
 CC to express the proteins. N and P can be used in the prevention, diagnosis  
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
 CC and AAH7789 represent sequences used in the exemplification of the  
 CC present invention.  
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
 CC missing at time of publication, meaning no sequences are present for  
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 1967 BP; 504 A; 438 C; 500 G; 524 T; 1 other;

Query Match 27.7%; Score 355.4; DB 22; Length 1967;

Best Local Similarity 66.5%; Pred. NO. 1.4e-88; Mismatches 256; Indels 0; Gaps 0;

Matches 509; Conservative 0; Mismatches 256; Indels 0; Gaps 0;

QY 159 aagacgtatcgcacagcttccagtgagcagctgaaagaaatgcccacaaataagaa 218  
 DB 116 acgcgcctcctcctcgcgcgtgagtcgcagcagcttgaccgcagaaagaaagcc 175  
 QY 219 gcaagcgttgcagaaatagaaagcaagcagcctgcagagttcggagagacttgaga 278  
 DB 176 gcgcgcctcctcgcagactcgcgcgcagcagctgcgcgttgagagagagacttgaga 235  
 QY 279 agagagctgctgcagagttgaaagccatctcaataaagtgtcttctgtagct 338  
 DB 236 aagaccctcagcggagagtcgggaaacgctattctcaagctaatggattgttgca 295  
 QY 339 gatgagaggagccgtcacacgcgtctacccacgcgcgtgatcaatgtcacagtgacaag 398  
 DB 296 gaagaaagaaagattacactgtttacaccccaacacacatctcacttgagacc 355  
 QY 399 atgtgtgacattcacaagtgtgaagtgtgtatcttagggcagaaccttccagagctcc 458  
 DB 356 atgtgtgacataaagatgtgaagtgctcatctctggaagatcatcatatgacact 415  
 QY 459 aaccagacatgactctgttctcagtgatgcaaaagccagttccctcccccagctc 518  
 DB 416 aacaaagctcagcgcgtcgtctttagtgcataaagccgttcgcctccgcagcttg 475  
 QY 519 gtgaacataacaaagattgtgtacccagatgagcttcaagaatccctgacatgga 578  
 DB 476 ggaacattataaagagtggtctacagacatagagatttcttcatcttgccatgga 535  
 QY 579 gatctaagcggatggcaaaaggggtgtgtgtcttaacggcgggtgtgacgcgtgcgg 638  
 DB 536 gatttctcgggtgtggaagcaaggtgtctcctctcctaagcgtctcccaagcttgcg 595  
 QY 639 gccatcagcgaactccacaaagagcagagcgtgggaaccttccacgaagcgtgtgac 698  
 DB 596 gccatcagcgaactccacaaagagcagagcgtgggaaccttccacgaagcgtgtgac 655  
 QY 699 aagtgctagcgtcacaacgggaagagtcgttctccttgggggagcgaatcccat 758  
 DB 656 tctgtgctaatacagaactcgaatgtgccttcttctgtcctcggggcctcttctgctag 715  
 QY 759 aagaagggagcgacacacgcagagaaagcgcacatgtcttgcagactgttccatcatc 818  
 DB 716 aagaagggagcgacacacgcagagaaagcgcacatgtcttgcagactgttccatcatc 775  
 QY 819 ccttgcctgcctcatcgtgtgttcttctgttgaagcacttccaaagagtaagcgctg 878  
 DB 776 ccttgcctgcctcatcgtgtgttcttctgttgaagcacttccaaagagtaagcgctg 835  
 QY 879 ctgaactatcttggagcggagccttaactgaagagcactctaa 923  
 DB 836 ctgcagaaagtcttggcagaagaccatctgactggaagagcgttga 880

RESULT 4  
 ID AAT73564  
 AC AAT73564 standard; cDNA; 2053 BP.  
 XX  
 AC AAT73564;  
 XX  
 DT 28-SEP-1997 (first entry)  
 XX  
 DE Uracil DNA glycosylase UNG2 cDNA.  
 XX  
 KW Uracil DNA glycosylase; UNG2; mutagenesis; DNA sequencing;  
 XX DNA modification; cell killing; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 71..1  
 FT /tag= a  
 FT mRNA 71..202  
 FT /tag= b  
 FT /note= "Claim 16"  
 FT mRNA 121..130  
 FT /tag= c  
 FT /note= "Claim 1"  
 XX  
 PD WO9725416-A2.  
 XX  
 PN 17-JUL-1997.  
 XX  
 PF 09-JAN-1997; 97NO-GB00057.  
 XX  
 PR 09-JAN-1996; 96GB-0000384.  
 XX  
 PA (DZIE/) DZIEGLEMSKA H E.  
 XX (NYFO-) NYFOTER AS.  
 XX  
 PI Kavli B, Krokan HE, Mol CD, Slupphaug G, Tainer JA;  
 XX  
 DR WPI: 1997-372857/34.  
 XX P-PSDB: AAM21814.  
 PT DNA glycosylase capable of releasing cytosine, thymine or uracil  
 PS bases from DNA - useful in vitro and/or vivo mutagenesis systems  
 XX to remove contaminating DNA prior to PCR amplification  
 XX  
 PS Claim 1: Page 47-48; 60pp; English.  
 XX  
 CC A cDNA clone (AAT73564) codes for a new form (UNG2) (AAM21814) of  
 CC uracil DNA glycosylase that is capable of releasing uracil bases  
 CC from single and/or double stranded DNA. It was isolated from a  
 CC human NT2 neuronal precursor cell cDNA library and represents an  
 CC alternatively spliced form of the UNG gene (see also AAT73567) that  
 CC arises by splicing of a previously unrecognized exon (exon 1A) into  
 CC a consensus splice site after codon 35 in exon 1B (previously  
 CC designated exon 1). UNG2 polypeptide differs from UNG1 in the  
 CC presence of a 44-amino acid N-terminal presequence (see also  
 CC AAM21817). Mutagenesis of UNG2 cDNA provides nucleic acids that  
 CC encode cytosine and thymine DNA glycosylases (see also AAM21815-16).  
 CC Recombinant DNA glycosylases can be expressed in host cells for use  
 CC in mutagenesis, to remove contaminating DNA prior to PCR, in DNA  
 CC modification and in cell killing.  
 XX  
 SO Sequence 2053 BP; 488 A; 486 C; 538 G; 541 T; 0 other;

Query Match 27.7%; Score 355.4; DB 18; Length 2053;  
 Best Local Similarity 66.5%; Pred. NO. 1.4e-88;  
 Matches 509; Conservative 0; Mismatches 256; Indels 0; Gaps 0;

QY 159 aagacgtatcgcacagcttccagtgagcagctgaaagaaatgcccacaaataagaa 218  
 DB 248 acgcgcctcctcctcgcgcgtgagtcgcagcagcttgaccgcagaaagaaagcc 307







```

XX 15-FEB-2001.
XX
XX 03-AUG-2000: 2000MO-US21176.
XX
XX 04-AUG-1999: 99US-0368382.
XX
XX (ABBO ) ABBOTT LAB.
XX
XX Chovan LE, Hessler PE, Reich KA;
XX
XX WPI: 2001-147511/15.
XX
XX P-PSDB; AAB8498.
XX
XX Essential bacterial genes from Haemophilus influenzae and methods for
XX identifying 'essential' genes that may be potential therapeutic targets
XX
XX
XX Disclosure: Page 52-53; 185pp; English.
XX
XX AAF94345 to AAF94409 represent essential bacterial genes from
XX Haemophilus influenzae, which encode the proteins given in AAB8492 to
XX AAB8556. The present invention also describes methods for identifying
XX essential bacterial genes (i.e. those essential to the survival of a
XX bacterium) using a transposition system. The methods are used to
XX identify essential genes from bacteria, especially H. influenzae (which
XX causes otitis media, meningitis and upper respiratory tract infections)
XX which may be used as targets for potential antimicrobial agents.
XX AAF94410 to AAF94416 represent PCR primers used in the exemplification
XX of the present invention.
XX
XX Sequence 657 BP; 213 A; 138 C; 130 G; 176 T; 0 other:
XX
Query Match          14.0%; Score 180; DB 22; Length 657;
Best Local Similarity 57.4%; Pred. No. 6.1e-40;
Matches 324; Conservative 0; Mismatches 240; Indels 0; Gaps 0;
XX
XX 348 agccgtcacacgctctaccacgagctgacagtgatgacagagatgtgtgac 407
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 88 agcgggaaacgattatcccccacagaagatgtattacgattcaataatagct 147
XX
XX 408 attcaagatgtgaagaagatgattctagagcagccttaccgcttcccaacga 467
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 148 ttgagagatgaaagaagtgatatttagtgacgattccattacgacaaacagcg 207
XX
XX 468 catggaactcgtttcaagtgcaaaagccagctcccccctccagctcgtgaata 527
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 208 caagcgtcgttttcttgaataaacctgaagtagcattcccccctccattataata 267
XX
XX 528 tacaagaatctgttaccgacatgtatgcttcaagatcccttgacatgagatc 587
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 268 tataaagaactcacacacagatatttcggtatttcaaatgacatcaaatgtatt 327
XX
XX 588 ggttggaacaaagggggtgctgttcaagcgtgtgacggtgagggccacag 647
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 328 aaatggcgacaaagggggtatgtctacttaacgtgttcaacgtggaagagatg 387
XX
XX 648 gcaaacctcccaagaagcagagctgtagaacttccacgagctgtgacatgagctg 707
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 388 gcaacattcacagcgaatttagtggtgaaaggtttacagataaagtattgacgtactc 447
XX
XX 708 agcgtcaacgggaagagtcgtttccctgtgtggtggtcacaagcccaagaagga 767
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 448 atgtaacatcgtgaaacacgtgttttcttacttggggcagtcacgacaaaaggg 507
XX
XX 768 ggcagacctgacaggaagacgtaccacatgcttgcgaagctgttaccatctcttct 827
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 508 caaatgctgacgcgactcgtcacttgtttaaagcgtccgcactcttcccgctgca 567
XX
XX 828 gctcatgttggtccttctgttgaagcacttcccaagcagcggtcgtcgaacta 887
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 568 gcaacatgagagttcttctgtgtcgtcattttcccaaaacaaatcatatttgaaagc 627

```

```

XX
XX 888 tctggagcagagcctataactg 911
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 628 cacggaataaacgcatagattg 651
XX
XX
XX RESULT 8
XX AAT42063
XX ID AAT42063 standard; DNA; 1830121 BP.
XX
XX AC AAT42063;
XX
XX 14-SEP-1999 (first entry)
XX
XX Haemophilus influenzae complete genome sequence.
XX
XX
XX Genome: bacterium; Haemophilus influenzae; computer readable medium;
XX expression modulating fragment; regulation; gene expression; vector;
XX organism; open reading frame; ORF; ds.
XX
XX Haemophilus influenzae.
XX
XX WO9633276-A1.
XX
XX 24-OCT-1996.
XX
XX 22-APR-1996; 96WO-US05320.
XX
XX 07-JUN-1995; 95US-0487429.
XX 21-APR-1995; 95US-0426787.
XX 07-JUN-1995; 95US-0476102.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (UYJO ) UNIV JOHNS HOPKINS.
XX
XX Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;
XX
XX WPI: 1996-485782/48.
XX
XX Haemophilus influenzae Rd genome recorded on computer readable
XX medium - useful for identifying commercially important nucleic acid
XX fragments by homology searching
XX
XX Claim 1; Page 77.2-77.1091; 1291pp; English.
XX
XX This sequence represents the complete genome sequence of the bacterium
XX Haemophilus influenzae strain Rd. The invention relates to a computer
XX readable medium (CRM) having recorded upon it the complete H. influenzae
XX nucleotide sequence (1), a representative fragment of (1) or a nucleotide
XX sequence at least 9% identical to (1). By providing the full-length
XX genomic sequence in a computer readable form, it is possible to identify
XX commercially important nucleic acid fragments and expression modulating
XX fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
XX regulate the expression of a nucleic acid molecule. Vectors and altered
XX organisms comprising the predicted ORFs can be used to produce any of the
XX polypeptide fragments of the H. influenzae Rd genome.
XX
XX Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;
XX
Query Match          14.0%; Score 180; DB 17; Length 1830121;
Best Local Similarity 57.48%; Pred. No. 3.6e-38;
Matches 324; Conservative 0; Mismatches 240; Indels 0; Gaps 0;
XX
XX 348 agccgtcacacgctctaccacgagctgacagtgatgacagagatgtgtgac 407
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 18174 agcgggaaacgattatcccccacagaagatgtattatcagcatccaatatagcgt 18233
XX
XX 408 attcaagatgtgaagaagatgattctagagcagccttaccaggttcccaacga 467
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 18234 ttgagagatgaaagaagtgatatttagtgacgattccattatcagcacaacagcg 18293
XX
XX 468 catggaactcgtttcaagtgtcgaaagccagttcccccctcccccagtcgtgaaacata 527

```

```

Db 18294 cagcgctgcttcttcagtaaaacccgaagtagccatcccccctcccttaataata 18353
Qy 528 tacaagaattggtaccgacatgatggtcttcaagcatccctgacatgagactaagc 587
Db 18354 tataaagaactcacacaagatattcgggatttcaaatgcatcaaatggtattatgctc 18413
Qy 588 ggaatggcaaaaacaagggtgctgctcttcaacgcggtgctgaacgcgcgcgcgcacacag 647
Db 18414 aaatggcagacaagaagggtatgcttacttaacacgttcttaccgtggaacgaggtatg 18473
Qy 648 gccaaatcccaagaagacagagctggagagaccttcaacgcgtgtgatcaagtgcgtg 707
Db 18474 gccaatcccaacgcacaaattaggttggaaaggtttacagataaagttatgctgacttc 18533
Qy 708 agatgcaacggggaaggagctgttctctgttgggggcgtacatcgcccatagaagaagg 767
Db 18534 aatgaacaatcgtagaaaactcgtggtttcttactcttggggcagctcgcacaaaaaagg 18593
Qy 768 ggcacacatcgacaggaacgltcacacatgcttgcgaagctgttccatccatctcttgc 827
Db 18594 caaatgttgcgcgcactcgtcaacctgttttaacggtccgcacatccctcccggtgtca 18653
Qy 828 gccatctgtgtggttctctgttgaagcacttctcaaggtcaaggcgtgctgaaacta 887
Db 18654 gccacatcgaggttcttctgttgcgtcatttcccaaaaacaattcataatttgaaagc 18713
Qy 888 tctggagcggagcctataactgg 911
Db 18714 cagcgaataaaacccgatagattgg 18737

RESULT 9
AAAF21609
ID AAF21609 standard; DNA: 349980 BP.
XX
AC AAF21609;
XX
DT 13-MAR-2001 (first entry)
XX
DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:110.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
diagnosis; antigen; detection; infection; gene therapy; antibacterial;
ds.
XX
OS Neisseria meningitidis.
XX
PN WO200066791-A1.
XX
PD 09-NOV-2000.
XX
PF 08-MAR-2000; 2000WO-US05928.
XX
PR 30-APR-1999; 99US-0132068.
PR 08-OCT-1999; 99WO-US23573.
PR 28-FEB-2000; 2000GB-0004695.
XX
PA (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Piza M, Hickey E, Peterson J, Tettein H, Venter JC, Masignani V,
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
PI Frazer CM, Grandi G;
XX
DR WPI: 2000-647603/62.
XX
PT Neisseria meningitidis B full length genome sequence and open reading
frames are used to detect, treat and prevent Neisserial infections -
XX
PS Claim 7: Appendix A: 692pp; English.
XX
CC The present invention describes the full length genome of

```

```

CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC to AAF21613 represent fragments of the NMB genomic sequence, as the
CC sequence was too long to go in a record on its own it was split into 8
CC sequences which overlap each other at the beginning and end of each
CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC AAF21606 represent PCR primers which are used in the exemplification of
CC the present invention. The NMB genome and fragments from it have
CC antibacterial activity, and can be used in vaccines and gene therapy.
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC proteins can be used in compositions for treating or preventing infection
CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC presence of Neisserial bacteria or of antibodies raised to Neisserial
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
CC used.
XX
SQ Sequence 349980 BP; 81351 A; 86755 C; 95584 G; 86290 T; 0 other;

```

```

Query Match 13.4%; Score 172.2; DB 21; Length 349980;
Best Local Similarity 56.7%; Pred. No. 2.3e-36;
Matches 318; Conservative 0; Mismatches 243; Indels 0; Gaps 0;

```

```

Qy 352 gtcaaccgtctaccacccggtcatcaatgatagttgcagagatgtgtacttc 411
Db 325875 gacaaatcatctatcccgcgcgagatgtgttcaacgacattccgctcagcgctg 325934
Qy 412 aagatgtgaagatgatactctagccagacaccttaccacagctgtcccaacaacatg 471
Db 325935 accggttcaaaagccgtcatctcgtgaagaatccgtatcaaggcgacggcagcgagc 325994
Qy 472 gactctgttaagtgtgcaaaagccagttcccccctcccccagctcgtgaacatata 531
Db 325995 gtttgcatcttccgtcgcgacaggtatccgcataccgcgcgtcttactcaatatcata 326054
Qy 532 aagaatgtgtaccgaatgatgatggtctaaagcatccctgtgacatgagatcagcgat 591
Db 326055 aggaagtgtgaacacccgacatcgaaggttccatcccgccagcgctgtctgacagcgt 326114
Qy 592 gggcaaaaacagggtgctgctgtcttcaacgcgtgtgcacgcgtgcgggccaataagcaca 651
Db 326115 gggcgagacggggcgtatgtcttctgaacaaggttttgacgggtgcgcaggaagcgcg 326174
Qy 652 actccacaagagacagagctggtggaagaccttaccagaagctgtgataagtgtgagcg 711
Db 326175 atgcacagccctttagctgtggaacgctttaccgataccgttatcaaggcagcttcgca 326234
Qy 712 tcaaccgggaagagatgcttctctgtgtggtggtcatataagcccaataaagaaggacga 771
Db 326235 caacccgaaagacactgtcttcaatgtgtggtggtgtgcacaaacaaaaggagcg 326294
Qy 772 ccacgcagagaaagcgtccacatgtcttgcagaagctgtcatccatcccttgcctc 831
Db 326295 tgatagacagtcaaaatcattgatattgacgcgcacgcacatccgtccctcgtgcgcat 326354
Qy 832 atcgtgtgttctctgttgttaagcacttccaaaggtcaaggcgtgcgtcaaatatctg 891
Db 326355 atcgggttttctgcgtgcgcacatttccacagsgaaacagattattgagccgcgacg 326414
Qy 892 ggcagagacctaataactgga 912
Db 326415 gtatcgatccgataaactgga 326435

```

```

RESULT 10
AAAF21610
ID AAF21610 standard; DNA: 349980 BP.

```

```
XX AAF21610;
AC
XX 13-MAR-2001 (first entry)
DT
XX
XX Neisseria meningitidis B nucleotide sequence SEQ ID NO:111.
DE
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
KM diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KW ds.
XX
XX Neisseria meningitidis.
OS
XX WO200066791-A1.
XX
XX 09-NOV-2000.
PD
XX
XX 08-MAR-2000; 2000MO-US05928.
PF
XX
XX 30-APR-1999; 99US-0132068.
PR 08-OCT-1999; 99MO-US23573.
PR 28-FEB-2000; 2000GB-0004695.
XX
XX (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
XX Pizsa M, Hickey E, Peterson J, Tettein H, Venter JC, Masigiani V;
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
PI Frazer CM, Grandi G;
XX
XX WPI: 2000-647603/62.
XX
XX Neisseria meningitidis B full length genome sequence and open reading
PT frames are used to detect, treat and prevent Neisserial infections -
XX
XX Claim 7: Appendix A: 692pp; English.
PS
XX
XX The present invention describes the full length genome of
CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC to AAF21613 represent fragments of the NMB genomic sequence, as the
CC sequence was too long to go in a record on its own it was split into 8
CC sequences which overlap each other at the beginning and end of each
CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC AAF21606 represent PCR primers which are used in the exemplification of
CC the present invention. The NMB genome and fragments from it have
CC antibacterial activity, and can be used in vaccines and gene therapy.
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC proteins can be used in compositions for treating or preventing infection
CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC presence of Neisserial bacteria or of antibodies raised to Neisserial
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
XX used.
XX
XX Sequence 349980 BP; 86771 A; 92803 C; 86340 G; 84066 T; 0 other;
SQ
```

```
Query Match 13.4%; Score 172.2; DB 21; Length 349980;
Best Local Similarity 56.7%; Pred. No. 2.3e-36;
Matches 318; Conservative 0; Mismatches 243; Indels 0; Gaps 0;
```

```
QY 352 gtaacacgctctaccacgctgcatcaagtgtacacatttgacaggaatgtgtacattc 411
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 25875 gacaaatcatctaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 25934
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
QY 412 aagatgtgaagtgtatctagtcagcagccctaccacgcgtcccaacaaacacatg 471
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
Db 25935 accggttcaaaagccgtatcttcgagacaagatcccgatcaaggcgagggcagcgacg 25994
QY 472 gactctgttcaagtgtgcaaaagccagttcccccctcccccagctcgtgacatataca 531
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 25995 gtttgcatcttccgtccgcgaaggatccgcacatccgcgcgtcttactacatatcata 26054
QY 532 aagattgtgtacagacatgatgtcttcaagacatccctcgagacatggagatcaagcgat 591
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 26055 aggaattggaaccgacacatcgaaagccttccatcccgccgacagcgtctctgacagcgt 26114
QY 592 gggcaaaacaaagggtgtctgtcttcaacgcgtgtcgtacccgtggtgcccataagagcca 651
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 26115 gggcggagcagggcgtatgtcttcttgacaacagcgttttgacggtgtgcgaagagcgc 26174
QY 652 actcccaaaagacagagcgtgtggaagaccttccacagcgtgtgatalcaagtgtcgtacg 711
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 26175 attcgacagccctttagcgtgtggaacgcttaccagcgttatacagcagcttcgcga 26234
QY 712 tcaacgggaaaggaggtgtcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 771
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 26235 cacacggcaagcactgtcttccatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 26294
QY 772 ccacgacaggaagacgtacacatgtcttcaagctgttcatcattcccttctgtctgtc 831
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 26295 tgatagacagtcacaaatcatttgatgtgacgcacgcacgcacgcacgcacgcacgcac 26354
QY 832 atcgtgggttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 891
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 26355 atcgcggttttctgtcgtcgcgcacatttccacagcacaacagatatttgccgcgacg 26414
QY 892 ggacggagcctataactgga 912
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 26415 gtatcgatccgataaactgga 26435
```

## RESULT 11

AAAX13540/C  
ID AAX13540 standard; DNA; 2469 BP.

AC AAX13540;

DT 19-MAR-1999 (first entry)

DE Enterococcus faecalis genome contig SEQ ID NO:603.

KW Enterococcus faecalis; contig; detection; Enterococcal infection;

KM vaccine; attenuation; computer readable medium; ds.

OS Enterococcus faecalis.

PN W09850555-A2.

PD 12-NOV-1998.

PF 04-MAY-1998; 98WO-US08985.

PR 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.

PR 16-MAY-1997; 97US-0046655.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Dillon PJ, Kunsch CA;

WPI: 1999-045171/04.

New isolated Enterococcus faecalis polynucleotides and polypeptides  
- used to develop products for the detection of Enterococcus and for  
use in vaccines for prevention or attenuation of Enterococcus  
infection.

Claim 1; Page 1853-1855; 2084pp; English.

CC A computer readable medium has been developed which has recorded on it  
 CC 982 nucleotide sequences isolated from the *Enterococcus faecalis* genome.  
 CC AAX12938 to AAX13919 represent these nucleotide sequences which are  
 CC primary nucleotide sequences, also known as contigs. The computer-based  
 CC system can identify fragments of the *Enterococcus faecalis* genome with  
 CC commercial importance. The products can be used to detect the presence  
 CC of *Enterococcus faecalis* in samples. They can also be used for  
 CC diagnosing *Enterococcal* infection in an animal and monitoring  
 CC progression of disease, and for identifying agents which can be used to  
 CC modulate the growth or pathogenicity of *Enterococcus faecalis*, or  
 CC another related organism, in vivo or in vitro. In particular the  
 CC polypeptides encoded by the *Enterococcus faecalis* nucleotide sequences  
 CC can be used in vaccines to prevent or attenuate an *Enterococcal*  
 CC infection.

XX Sequence 2469 BP; 771 A; 471 C; 483 G; 738 T; 6 other;

Query Match 12.5%; Score 160.8; DB 20; Length 2469;  
 Best Local Similarity 54.2%; Pred. No. 2,7e-34;  
 Matches 349; Conservative 0; Mismatches 292; Indels 3; Gaps 1;

QY 268 agacttgagaagagagctgctgctgagcttgaagaagcattcaacaattgagt 327  
 DB 999 ATAGTTGGCAAAACATTATTCTGCTCAATTTGAAAAACGTATTATCAAGATTGGCGG 940  
 QY 328 cctttagctgtagtagagagagccgtacacacgtctaccacagcgctgatacga 387  
 DB 939 AATTTTGAAGAAAAAGATATCAGACACAAAGATTATCCAGATATATCATTTATTTT 880  
 QY 388 gtccgaagaagtagtgatgacattcaagaatgtgaagaatgagtctgagcagaccc 447  
 DB 879 CAGGTYAGACTAACCCCTTTTGAAGAGATTAAGATTTTGGGCAAGATCAT 820  
 QY 448 accacggtcccaaccaacacacatgactcgttctcagtgagcaaaagccagctcc 507  
 DB 819 ATCATGAGCCCAATCAACGACACAGGACTAAGTTCTCTGTCACACAGGTGTAAGTGC 760  
 QY 508 ccccaagctcgtgacacatacaagaattgtgtacagacatgagtgcctcaagc 567  
 DB 759 CGCATCTCTTGGCAATATTACAAAGATTACAAAGCTGA--TTTAGGGTACCAACAG 703  
 QY 568 ctgagacatgagatctgaagagatgggcaaaagaggtgctgctcttaacgagtc 627  
 DB 702 TCAACCATGGCTTTTGAAGAGCTGGGCCAAACAGCGTGTATTATTAATATCTGT 643  
 QY 628 tgaacgtgagggccatcagcagcaaccccaacaagagacagagctgggagac 687  
 DB 642 TGAACGTTGGGCGACGTCAGACCTATTCATCGACGTAAGGGTGGGACCACTGACAG 583  
 QY 688 acgctgtgataagtgctgtagcgctcaacggggaaggagcttccctgtgtgggct 747  
 DB 582 ATGTCATTATCGAAATGATGAGCGTGAATAAACTGTGTTTCAATTTATGAGGAG 523  
 QY 748 cataagccctaaagaagggcgacccatcgacaggaagcgtccacatgctctgaagct 807  
 DB 522 GTCCCTGCTCAAGAAAAATTAAATGATTATAGCATCGCATGTGATTAATCAATCGC 463  
 QY 808 ttcataatcctcttctgctcattcgttggttccctgttgtaagcaacttcccaag 867  
 DB 462 CACACCCCAAGTCCATTGTCAGCACATCGTGTTTGGCTACGTCGCTTTTCTCAAG 403  
 QY 868 ctaagggctgctgaactactctggaagcagcctataaactgg 911  
 DB 402 CGAATGCCGCTCTAGACGTTTAGGGGAACGCCGATTTGATTGG 359

RESULT 12  
 AAX27774  
 ID AAX27774 standard; DNA; 681 BP.  
 XX  
 AC AAX27774;  
 XX

DT 03-JUN-1999 (first entry)  
 XX  
 DE B. pallidus *ureacil* DNA glycosylase enzyme (Bpa UDG) encoding DNA.  
 XX  
 XX *Uracil* DNA glycosylase: Bpa UDG; enzyme: catalyse; hydrolysis;  
 KW nucleic acid amplification; thermostable; ss.  
 KW *Bacillus pallidus*.

OS  
 XX  
 PN US5888795-A.  
 PD 30-MAR-1999.

XX 09-SEP-1997; 97US-0926055.

XX 09-SEP-1997; 97US-0926055.

XX (BECT ) BECTON DICKINSON & CO.

XX Hamilton PT;

XX WPT; 1999-243212/20.

XX P-PSDB; AAY01335.

XX Nucleic acid encoding *ureacil* DNA glycosylase from *Bacillus pallidus*  
 XX  
 PS Claim 1; Columns 7-12; 7pp; English.

CC This DNA encodes a novel *ureacil* DNA glycosylase enzyme (Bpa UDG) of  
 CC *Bacillus pallidus*. Host cells transformed with the vector comprising the  
 CC Bpa UDG gene can be used for the recombinant expression of the protein.  
 CC The enzyme selectively catalyses hydrolysis of the N-glycosylation link  
 CC between *ureacil* and deoxyribose, creating an a basic site that is  
 CC susceptible to acid or alkaline hydrolysis. UDG is used to inactivate  
 CC contaminating amplicons (containing *ureacil*) in nucleic acid amplification  
 CC mixtures, particularly at high temperature. Antibodies raised against UDG  
 CC can be used to purify or detect UDG (in usual immunoassays), also to  
 CC inactivate it, e.g. as a replacement for, or in combination with, heating  
 CC or use of *ureacil* glycosylase inhibitory protein, for control of UDG  
 CC activity. The Bpa UDG, which has similar kinetics to the enzyme from  
 CC *E. coli* is thermostable, thereby allowing decomposition of contaminating  
 CC amplicons at higher temperatures, e.g. 45-75 deg. C than is possible with  
 CC known mesophilic enzymes. This reduces the size of temperature changes  
 CC between the various stages of the decontamination process.

XX Sequence 681 BP; 210 A; 128 C; 159 G; 184 T; 0 other;

Query Match 12.4%; Score 159.6; DB 20; Length 681;  
 Best Local Similarity 53.8%; Pred. No. 3e-34;  
 Matches 352; Conservative 0; Mismatches 299; Indels 3; Gaps 1;

QY 260 tttagagagagacttgagaagagagctgctgacagattggaagcaccatactcaaca 319  
 DB 12 tttaaaaatgagatggtggggagctgtggaagaagatlttgaagaagcaccatacaaga 71  
 QY 320 atgatactcccttgtagctgtagagagagcggtacacacgctctacccacgctgata 379  
 DB 72 gctgcgggaatttttgaagcagaatatagccacatacagattatccgagatagtaca 131  
 QY 380 agtataagttgcacaagaatgtagtgcacattcaagaatgtaagaatgatttcaagcca 439  
 DB 132 cattttaagcttctgattacacgcttatagaagaatgaagttgctatttaagcca 191  
 QY 440 ggaaccttaccaggtcccaaccaagacatgagctgttctcagtggtgcaaaagcagct 499  
 DB 192 agatcgtatcatgaccacaatcaacacatgagtttaagcttttcgttaagccggggat 251  
 QY 500 tccccctcccccagctcgttgaacatacaagaatgtgtacgcagacattgagctt 559  
 DB 252 tgcacagcgcgcgtctctgtaaaatatttcatcgagcttgagaaga---ttcggctg 308  
 QY 560 caagatcctcgagacatgagatctaagcggatggtgcaaaaagaagggtgctgctgctaa 619





RESULT 14  
AAC36561  
ID AAC36561 standard; DNA; 1348 BP.  
XX  
AC AAC36561;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 14258.  
XX  
KW Hybridization assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PE 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123848.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148371.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148365.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.

```
PR 20-AUG-1999: 99US-0149929.
PR 23-AUG-1999: 99US-0149902.
PR 23-AUG-1999: 99US-0149930.
PR 25-AUG-1999: 99US-0150566.
PR 25-AUG-1999: 99US-0150884.
PR 27-AUG-1999: 99US-0151065.
PR 27-AUG-1999: 99US-0151066.
PR 27-AUG-1999: 99US-0151080.
PR 30-AUG-1999: 99US-0151303.
PR 31-AUG-1999: 99US-0151438.
PR 01-SEP-1999: 99US-0151930.
PR 07-SEP-1999: 99US-0152363.
PR 10-SEP-1999: 99US-0153070.
PR 13-SEP-1999: 99US-0153758.
PR 15-SEP-1999: 99US-0154018.
PR 16-SEP-1999: 99US-0154039.
PR 20-SEP-1999: 99US-0154779.
PR 22-SEP-1999: 99US-0155139.
PR 23-SEP-1999: 99US-0155486.
PR 24-SEP-1999: 99US-0155659.
PR 28-SEP-1999: 99US-0156458.
PR 29-SEP-1999: 99US-0156596.
PR 04-OCT-1999: 99US-0157117.
PR 05-OCT-1999: 99US-0157753.
PR 06-OCT-1999: 99US-0157865.
PR 07-OCT-1999: 99US-0158029.
PR 08-OCT-1999: 99US-0158232.
PR 12-OCT-1999: 99US-0158369.
PR 13-OCT-1999: 99US-0159293.
PR 13-OCT-1999: 99US-0159294.
PR 13-OCT-1999: 99US-0159295.
PR 14-OCT-1999: 99US-0159329.
PR 14-OCT-1999: 99US-0159330.
PR 14-OCT-1999: 99US-0159331.
PR 14-OCT-1999: 99US-0159637.
PR 14-OCT-1999: 99US-0159638.
PR 18-OCT-1999: 99US-0159584.
PR 21-OCT-1999: 99US-0160741.
PR 21-OCT-1999: 99US-0160767.
PR 21-OCT-1999: 99US-0160768.
PR 21-OCT-1999: 99US-0160770.
PR 21-OCT-1999: 99US-0160814.
PR 21-OCT-1999: 99US-0160815.
PR 22-OCT-1999: 99US-0160980.
PR 22-OCT-1999: 99US-0160981.
PR 22-OCT-1999: 99US-0160989.
PR 25-OCT-1999: 99US-0161404.
PR 25-OCT-1999: 99US-0161405.
PR 25-OCT-1999: 99US-0161406.
PR 26-OCT-1999: 99US-0161359.
PR 26-OCT-1999: 99US-0161360.
PR 26-OCT-1999: 99US-0161361.
PR 28-OCT-1999: 99US-0161920.
PR 28-OCT-1999: 99US-0161922.
PR 28-OCT-1999: 99US-0161993.
PR 29-OCT-1999: 99US-0162142.
PR 29-OCT-1999: 99US-0162142.
```

Query Match 11.1%; Score 142.2; DB 21; Length 1348;  
Best Local Similarity 54.3%; Pred. No. 3.1e-29;  
Matches 310; Conservative 0; Mismatches 258; Indels 3; Gaps 1;

```
QY 341 tggagagagcgcgtcacacgcgtctaccaccgcggtgtatcaagtgtacagttcgacagagat 400
DB 483 tgcacgtlaaagaagccctctgtattatccaccgcgacactgtatttcaatgtcttataac 542
QY 401 gtgtgacattcaagatgtgaagtagtattctagccagacccttaccagagttcccaa 460
DB 543 aactccttttgatcgaggttaagactgtcatattacgcagagatccctatatcatggaacctg 602
QY 461 ccaagacatggaactctgtttcagttgtgcaaaagccagttcccccctcccccagttcgt 520
DB 603 tcaagcatgtggtttgtctctctctgtactgtagagagaagaaagcttccctctagtctgt 662
```

```
QY 521 gaacatatacaagaattgtgtacccagacattatggtccttaagcattccgtgacatgagaga 580
DB 663 gaacattcttaagagattcatcaagaatggtgtctgttccatcccaac---gtcaggttaa 719
QY 581 tctaagcggatggtgcaaaaaggggtgtcgtgtcttaagcgggtgtctgacggtgtgggc 640
DB 720 tctacagaatggtgctgtgcaagggtgtgttaactcccgaaatgctgttcttcaagtaaggag 779
QY 641 ccataagggccaactcccaagaagcagaggtgtggtgagaccttaccagacgtctgtatcaa 700
DB 780 taacagccttaattcatatcatgcaagaaggaatggtggaacaattacacgtatgtatttca 839
QY 701 gtgtcgtgagcgttcaaccgggaaggagtgcttcttcctgtgtgtggtgtcacaagccataa 760
DB 840 aagttatccacagcaagaaggaagggtgtgttcttctctcctgtgggaagatacgtcaaga 899
QY 761 gaaggagcagaccattcgacaggaagaacgttaacatgtcttgcaagctgttatcatctcc 820
DB 900 gaattccaagttagatagatgagactaaacatacatatactacacagcgtcatccatctgg 959
QY 821 ttgtctgtcatcgttgggttctctgtgtgttaagcacttccaaaggttaacggtctgtct 880
DB 960 ttgtcggcgaatagaggtcttcttcgtactcaggtcaggtcattctctcgtcgcaaacagctact 1019
QY 881 gaactatctggtgacggaacctataactgg 911
DB 1020 cgaggaaatggtggtatcttcccatagactgg 1050
```

## RESULT 15

ID ABA03041 standard; DNA: 2944528 BP.

AC ABA03041;

DT 05-FEB-2002 (first entry)

DE *Listeria monocytogenes* EGD-e genome sequence.

KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;

KW vitamin B12; bacterial infection; disease; ds.

OS *Listeria monocytogenes*.

XX WO200177335-A2.

XX 18-OCT-2001.

PF 11-APR-2001: 2001MO-FR01118.

XX 11-APR-2000: 2000FR-0004629.

PA (INSP ) INSR PASTEUR.

XX Buchrieser C, Frangoul L, Couve E, Rusnlok C, Fsihi H, Dehoux P;

PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;

PI Daniels J, Goebel W, Krefte J, Kuhn M, Ng E, Vazquez-Boland JA;

PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend L;

PI Chakraborty T, Dommann E, Hain T, Berche P, Charbit A, Durant L;

PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;

PI Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;

PI Rose M, Voss H;

XX MPI: 2002-010914/01.

XX Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment

PT and prevention of *Listeria* and related bacterial infections, and

PT related polypeptides -

XX Claim 1, SEQ ID No 1; 192pp; French.

CC The present sequence is the genome sequence of *Listeria monocytogenes* EGD-e. This sequence and fragments of this sequence are useful for

CC selecting probes and primers for detecting genes in L. monocytogenes and  
CC related organisms, and to study genetic polymorphisms and other genomes.  
CC Proteins (AB847297-AB850149) expressed from the present sequence are  
CC useful for raising specific antibodies, identification of L.  
CC monocytogenes and related organisms, and for biosynthesis and  
CC biodegradation, especially biosynthesis of Vitamin B12. This sequence and  
CC proteins encoded by it are also useful for selecting compounds that  
CC regulate gene expression and cell replication and modulate L.  
CC monocytogenes-related diseases. In addition, this sequence and proteins  
CC encoded by it are useful in pharmaceutical and vaccines compositions for  
CC the treatment or prevention of infections by L. monocytogenes and related  
CC organisms.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX

Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;

Query Match 10.8%; Score 138; DB 24; Length 2944528;  
Best Local Similarity 51.8%; Pred. No. 2.4e-26;

Matches 337; Conservative 0; Mismatches 310; Indels 3; Gaps 1;

QY 262 tcggagagacttggagaagagagctgctcagagatttgaagaagccatctcaacaat 321  
Db 414096 tggtaaaacatgggaagaattttaaacagaagaacaacagcctattatagaat 414155  
QY 322 tgaatgctcttgatcgaagagagccgtacacgcgtacccacgcgcatcaag 381  
Db 414156 taatggaagcagcgaagatcgagagagaaggtaatggtatccgtcagagaaata 414215  
QY 382 tgtacagctcgacaagagatgtgacatccaagatggaagtgaatgattctagccagg 441  
Db 414216 tgtttcgttttcgttcgtatgctgataatcaagtaagtcgtatttttgggaag 414275  
QY 442 acccttaccacggtcccaacagacatgactctgttcagtgctgcaaaagccaattc 501  
Db 414276 atccgtatccggtctcgagcaagcgatgcttaagcttccgtacaaaagaagctac 414335  
QY 502 cccctcccccagctctcgtacacataacaaagaattgtgtaccgacatgagcttca 561  
Db 414336 gaattccacctagctctcgaaatattataaagaatgaagacagacttagat--atcg 414392  
QY 562 agcatcctgacatgagatctaagcggatggcacaagaagggtgctgctcttaag 621  
Db 414393 aaccagacagaccacggtatctgtctaaatggtcagagcaaggcgtactactatgaca 414452  
QY 622 cgggtgctgacggtgcgggcccatcagggccaaccccaagagcagagcttggagacct 681  
Db 414453 cgggttgggagcgttgaagggaagaaagctggtcgagccacaaaactagctgggcaagct 414512  
QY 682 tcaacgacgctgtgatcaagttggtcagcgtcaacgggaaggagtcgtttccgtgt 741  
Db 414513 tcacagatcacgcttttagaagaactgataatgaacaaacacacgtctattttat 414572  
QY 742 ggggctcatcagcccatagaagagggcgacacatcgacaggaagcctacacatgctctgc 801  
Db 414573 ggggaaatactgtctataaagccgcgaagtggcatctacaacccgcgaacatctaat 414632  
QY 802 aagctgttcaatccatctcttctgtctcatcgttgggtctcttggttgtaagcacttct 861  
Db 414633 aagagatcacacatcacgcttgcgcgagctgagagcttttttggagaacacatttt 414692  
QY 862 ccaaggtctaagcgtgctgaaactatctggagcggagcctataaactgg 911  
Db 414693 cgaanaacgaatgcgtttcttgaagaacatgaaagaacacaaatgtgtg 414742

Search completed: August 22, 2002, 05:48:38  
Job time: 6843 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 22, 2002, 03:37:10 : Search time 83.01 Seconds  
(without alignments)  
3796.504 Million cell updates/sec

Title: US-09-758-017A-1

Perfect score: 1283

Sequence: 1 gacatccgctgcacaatg.....agaataatatatttta 1283

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents NA: \*  
1: /cgn2\_6/prodata/2/lna/5A.COMB.seq: \*  
2: /cgn2\_6/prodata/2/lna/5B.COMB.seq: \*  
3: /cgn2\_6/prodata/2/lna/6A.COMB.seq: \*  
4: /cgn2\_6/prodata/2/lna/6B.COMB.seq: \*  
5: /cgn2\_6/prodata/2/lna/PCITUS.COMB.seq: \*  
6: /cgn2\_6/prodata/2/lna/Backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	207.6	16.2	1532	4	US-09-651-656-36
2	107	8.3	35100	2	US-08-770-379-18
3	107	8.3	35100	4	US-08-757-669A-18
4	107	8.3	35100	4	US-09-230-371A-18
5	89	6.9	4403765	4	US-09-103-840A-2
6	89	6.9	4411529	4	US-09-103-840A-1
7	48.2	3.8	7218	1	US-08-232-463-14
8	46.4	3.6	7218	1	US-08-232-463-14
9	41.2	3.2	19124	2	US-08-487-826B-13
10	39.4	3.1	72604	4	US-09-268-992-7
11	37.4	2.9	1689	1	US-07-991-867B-41
12	37.4	2.9	1689	2	US-08-544-332-41
13	37.4	2.9	8457	1	US-07-991-867B-1
14	37.4	2.9	8457	2	US-08-544-332-1
15	37	2.9	1850	3	US-08-617-860B-32
16	37	2.9	4098	2	US-08-605-106-4
17	36.2	2.8	636	4	US-08-998-416-1137
18	36.2	2.8	837	4	US-08-998-416-288
19	36.2	2.8	1411	3	US-09-188-930-12
20	36	2.8	2223	1	US-08-257-073-4
21	35.8	2.8	319	1	US-07-593-657-14
22	35.4	2.8	678	1	US-07-991-867B-23
23	35.4	2.8	678	1	US-08-107-755A-23
24	35.4	2.8	678	2	US-08-544-332-23
25	35.4	2.8	4818	3	US-08-817-926-27
26	35.4	2.8	6768	1	US-08-107-755A-1
27	35.4	2.8	8457	1	US-07-991-867B-1

C	28	35.4	2.8	8457	2	US-08-544-332-1	Sequence 1, Appl
C	29	34.6	2.7	615	4	US-08-998-416-186	Sequence 186, App
C	30	34.6	2.7	6519	1	US-08-233-008A-7	Sequence 7, Appl
C	31	34.4	2.7	826	2	US-08-238-821B-51	Sequence 51, Appl
C	32	34.4	2.7	826	2	US-08-238-821B-61	Sequence 61, Appl
C	33	34.4	2.7	826	5	PCT-US95-05744-51	Sequence 51, Appl
C	34	34.4	2.7	826	5	PCT-US95-05744-61	Sequence 61, Appl
C	35	34.2	2.7	5181	1	US-08-257-073-10	Sequence 10, Appl
C	36	34.2	2.7	5703	4	US-09-280-590A-36	Sequence 36, Appl
C	37	34.2	2.7	6243	2	US-09-056-075-1	Sequence 1, Appl
C	38	34	2.7	3108	3	US-08-968-752B-3	Sequence 3, Appl
C	39	34	2.7	3114	4	US-09-107-149-18	Sequence 18, Appl
C	40	34	2.7	3222	3	US-08-968-752B-1	Sequence 1, Appl
C	41	34	2.7	3852	1	US-07-867-106-2	Sequence 2, Appl
C	42	34	2.7	9048	3	US-08-973-273-4	Sequence 4, Appl
C	43	34	2.7	11225	6	5182210-9	Patent No. 5182210
C	44	34	2.7	51952	3	US-08-947-823-1	Sequence 1, Appl
C	45	34	2.7	246240	2	US-08-724-394A-20	Sequence 20, Appl

#### ALIGNMENTS

RESULT 1									
US-09-651-656-36									
Sequence 36, Application US/09651656									
Patent No. 6340566									
GENERAL INFORMATION:									
APPLICANT: MCCUTCHEN-MALONEY, SANDRA									
TITLE OF INVENTION: DETECTION AND QUANTITATION OF SINGLE NUCLEOTIDE									
POLYMORPHISMS, DNA SEQUENCE VARIATIONS, DNA MUTATIONS,									
FILE REFERENCE: IL-10689									
CURRENT APPLICATION NUMBER: US/09/651,656									
CURRENT FILING DATE: 2000-08-29									
PRIOR APPLICATION NUMBER: 60/192,764									
PRIOR FILING DATE: 2000-03-28									
NUMBER OF SEQ ID NOS: 106									
SOFTWARE: PatentIn Ver. 2.1									
SEQ ID NO 36									
LENGTH: 1532									
TYPE: DNA									
ORGANISM: Escherichia coli									
US-09-651-656-36									
Query Match									
Best Local Similarity 16.2%; Score 207.6; DB 4; Length 1532;									
Matches 388; Conservative 0; Mismatches 279; Indels 3; Gaps 1;									
QY	258	ggttcggaagacttgagaagagactgctgcacagtttgaaagccatact--tc 314							
DB	536	gctaacgaatlaacactgcatgacgctgctgctgaagaagacacacactatttc 595							
QY	315	aaacaattgctccttgatgactgagagagacgcgtacacacgctcaccacgct 374							
DB	596	ataacacctcagacgcgcgcagcagcgacgctgcgcctacatctaccaccaa 655							
QY	375	gatacagtgtaacgcttgacaagagatgctgacattcaaatgtgaagaagatcta 434							
DB	656	aaagatgcttctaagcgttcgcgtcttaacagagltgggtgacgtlaaagtggatctc 715							
QY	435	ggcagacaccttaccacggttcccaacacgacacatgactctgctcagtgcaaaag 494							
DB	716	ggcagacatcttatacagcagcagcagcagcagcagcagcagcagcagcagcagc 775							
QY	495	ccagttccctcccccagctctcgtgaacatatacaagaattgctacgacatgat 554							
DB	776	ggatctcagatctctcgtcgtacattatgataatgataaagagctggaaatactatccg 835							
QY	555	ggcttcagatctctcgtacattatgataaagagctggaaatactatccg 614							

Db 836 ggcctaccgcccctaatcatggtatcttgaaagctggcgctgcaagcgcttcgtcta 895  
 QY 615 cttaacgagtgctgaacgctgagccatcaagcccaaccccaagcaagagctg 674  
 Db 896 ctcaatctgtgtgagctgacgagctgcaagctgcaatcccaagccagctgctg 955  
 QY 675 ggaacctcaccgagctgtgataagctgagcgctcaacccggaagagctgcttc 734  
 Db 956 gaaacctccacgataagagctgataagcctgataaccagacgcggaagcgctgttc 1015  
 QY 735 ctgttgaggagctcatcagccataaagaaggagcgacccctgacagaagcctaccat 794  
 Db 1016 tcttgaggagctcgcgcgcaaaagaaggcgatataagataagcaagccatcat 1075  
 QY 795 gctctgaagctgtcatcatcctctgtctgtcgtcgtgctgctgctgctgaag 854  
 Db 1076 gtaactgaagcaacgcagctcgcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1135  
 QY 855 caactctcaagctcaagcgctgctgaaactatctggaagcgagcgctataactgaga 914  
 Db 1136 catttgctgagcaaatcagctgctgaaacagctgagcgagacgcatgactgagatg 1195  
 QY 915 gcaactcaac 924  
 Db 1196 ccaagtattac 1205

## RESULT 2

US-08-770-379-18/c  
 ; Sequence 18, Application US/08770379  
 ; Patent No. 5849564

## GENERAL INFORMATION:

; APPLICANT: Chang, Yuan  
 ; APPLICANT: Bohenzky, Roy A.  
 ; APPLICANT: Russo, James J.  
 ; APPLICANT: Edelman, Isidore S.  
 ; APPLICANT: Moore, Patrick S.  
 ; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED  
 ; TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 20  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Cooper & Dunham LLP  
 ; STREET: 1185 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: U.S.A.  
 ; ZIP: 10036

## COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/770,379  
 ; FILING DATE:

## CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

; NAME: White, John P.  
 ; REGISTRATION NUMBER: 28,678  
 ; REFERENCE/DOCKET NUMBER: 52342  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 278-0400  
 ; TELEFAX: (212) 391-0525

## INFORMATION FOR SEQ ID NO: 18:

; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 35100 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; US-08-770-379-18

Query Match 8.3%; Score 107; DB 2; Length 35100;  
 Best Local Similarity 52.8%; Pred. No. 6.1e-22;  
 Matches 255; Conservative 0; Mismatches 225; Indels 3; Gaps 1;

QY 356 caccgttaccaccgagctatcaagtgtacagttgacagagatgtgtacattcaaga 415  
 Db 34128 CACAGGTATCCCCACAGAGACAAGCTGAGTGTGTCTCCACTGCTGCATCCAGGGA 34069  
 QY 416 tgtgaagtagtgaattcaagccagagacccttaccagctgcccacccaagacagact 475  
 Db 34068 TATTAAAGTGATATCTTAAGCCAGACCCGTACACAAAGG---CCAAGTACTGGCT 34012  
 QY 476 ctgttcaagtgtgaaagccagtlccccctcccccaagctcgtgtaacalataaaga 535  
 Db 34011 GGCGTTAGTGTGATCCGCAATGTCCAGTTCCACCAGTTTGAGAGCATCTTTAGGA 33952  
 QY 536 atgtgtaccgacattatgtgtcttcaagcaacctgagacatgagagatcaagcgatggc 595  
 Db 33951 GCTAGAGGCTTCCGCTCCCAATTTCACTACTCTTCCACGGGTGCTCCGACAGCTGGGC 33892  
 QY 596 aaaaagaagggtgctgctgcttaacgcggtgctgacccgtgagccatcaagccaaatc 655  
 Db 33891 TCGCAGAGGTGTGTGTACTTAAACACAGTTTGGACGCTGAGAGAGGGAGGGCGGCTC 33832  
 QY 656 ccacaagagacagagctggaagaccttaccgagcgtgtgataagtgctgaagctcaa 715  
 Db 33831 ACACGAGGAGCTTGGCTGGGATTGGTTACAGAGTTTCATCATCATAGTATATCTCAAA 33772  
 QY 716 ccggaagagagctgcttctcgtgtgtggtcctacagccataaagaaggagcgacat 775  
 Db 33771 GTTAGAATGCTGCTTTCTCTCTGTGGGGCCGACAGCCATTGACAGAACTCCGCTCAT 33712  
 QY 776 cgacaagaaagctcaccatgtcttgcagagctgtgcaagctgtcatcctcctgtcgtcatcg 835  
 Db 33711 AAACGACAGAAACACTGCTGTGTAGCGCCAGCATCATCCGCTGCTGTGG 33652  
 QY 836 tgg 838  
 Db 33651 TGG 33649

## RESULT 3

US-08-757-669A-18/c  
 ; Sequence 18, Application US/08757669A  
 ; Patent No. 6183751

## GENERAL INFORMATION:

; APPLICANT: Chang, Yuan  
 ; APPLICANT: Bohenzky, Roy A.  
 ; APPLICANT: Russo, James J.  
 ; APPLICANT: Edelman, Isidore S.  
 ; APPLICANT: Moore, Patrick S.  
 ; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS  
 ; TITLE OF INVENTION: SEQUENCES AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 20  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Cooper & Dunham LLP  
 ; STREET: 1185 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: U.S.A.  
 ; ZIP: 10036

## COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/757,669A  
 ; FILING DATE:

## CLASSIFICATION: 424

## ATTORNEY/AGENT INFORMATION:

; NAME: White, John P.  
 ; REGISTRATION NUMBER: 28,678







TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: PTZ9pt-Fls  
US-08-232-463-14

Query Match 3.8%; Score 48.2; DB 1; Length 7218;  
Best Local Similarity 9.9%; Pred. No. 0.0025;  
Matches 38; Conservative 181; Mismatches 164; Indels 0; Gaps 0;

QY 788 tcacacgtcttcgaacgtcttcacatctctctgtcgtccatcggtggtccttg 847  
DB 1128 yyy 1187  
QY 848 ttgtaagcactctccaagcctgaagcgtcgtgaactatctggaagcctataa 907  
DB 1188 yyy 1247  
QY 908 ctggaagcactcttaactcttaactgtccttaactgtttaagatgaacat 967  
DB 1248 yyy 1307  
QY 968 cacactatcttctacagcttctccaagtlcaacacatcttaacttcttct 1027  
DB 1308 yyy 1367  
QY 1028 ttggaatgactgtcgtcttctgtcgtttagataactaaacacttaccactgca 1087  
DB 1368 yyy 1427  
QY 1088 tcttgactactgtcagtaataacttcaacttgacaacaaatgtattataa 1147  
DB 1428 yyyyyyygtacaaattcttacttcttacttacttgcataagataatgacagtg 1487  
QY 1148 ttgatatctctgtatataa 1170  
DB 1488 atgcctacatgccgttttgaa 1510

## RESULT 8

US-08-232-463-14/C  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: PTZ9pt-Fls  
US-08-232-463-14

Query Match 3.6%; Score 46.4; DB 1; Length 7218;  
Best Local Similarity 4.7%; Pred. No. 0.00086;  
Matches 17; Conservative 196; Mismatches 147; Indels 0; Gaps 0;

QY 108 aatlaatgaagaacagccgaagaactgagtcctcaaatgtggaacaaagcgtca 167  
DB 1391 RRR 1332  
QY 168 tcgcacagcttcaagtgaagcagctgaaagaatgcgcaaaataagaagcagcct 227  
DB 1331 RRR 1272  
QY 228 gacaagattagaaagcagccctgcaggtctcggaagcattgagaagagcgtg 287  
DB 1271 RRR 1212  
QY 288 gctcagagttgaaagcactcaacaattgactgtcctttagttagtgaagag 347  
DB 1211 RRR 1152  
QY 348 agccgtcacacgctctccacgcgctgatacagtgtaacgttcagacagagatgtgac 407  
DB 1151 RRR 1092  
QY 408 attcaagatgtgaagtagtattctagccagagaccccttacacggtcccaacagca 467  
DB 1091 RRR 1032

## RESULT 9

US-08-487-826B-13/C  
Sequence 13, Application US/08487826B  
Patent No. 5993827  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chluis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660



```

1      FILING DATE: 30-JAN-1992
2      PRIOR APPLICATION DATA:
3      APPLICATION NUMBER: US 07/657,584
4      FILING DATE: 19-FEB-1991
5      ATTORNEY/AGENT INFORMATION:
6      NAME: Saliwanchik, David R.
7      REGISTRATION NUMBER: 31,794
8      REFERENCE/DOCKET NUMBER: F7114.C3
9      TELECOMMUNICATION INFORMATION:
10     TELEPHONE: 904-375-8100
11     TELEFAX: 904-372-5800
12     INFORMATION FOR SEQ ID NO: 41:
13     SEQUENCE CHARACTERISTICS:
14     LENGTH: 1689 base pairs
15     TYPE: nucleic acid
16     STRANDEDNESS: single
17     TOPOLOGY: linear
18     MOLECULE TYPE: DNA (genomic)
19     ;
20     US-07-991-867B-41

```

Query Match	2.98;	Score 37.4;	DB 1;	Length 1689;
Best Local Similarity	46.38;	Pred. No. 0.2;		
Matches 157;	Conservative 0;	Mismatches 181;	Indels 1;	Gaps 1.

[illegible]

```

1 RESULT 12
2 US-08-544-332-41
3 Sequence 41, Application US/08544332
4 Patent No. 5935777
5 GENERAL INFORMATION:
6 APPLICANT: Moyer, Richard W.
7 APPLICANT: Hall, Richard L.
8 APPLICANT: Gruidl, Michael E.
9 TITLE OF INVENTION: No. 5935777e1 Entomopoxvirus Expression System
10 NUMBER OF SEQUENCES: 77
11 CORRESPONDENCE ADDRESS:
12 ADDRESSEE: Gerard H. Bengen
13 STREET: 2421 N.W. 41st Street, Suite A-1
14 CITY: Gainesville
15 STATE: FL
16 COUNTRY: USA
17 ZIP: 32606
18 COMPUTER READABLE FORM:
19 MEDIUM TYPE: Floppy disk
20 COMPUTER: IBM PC compatible
21 OPERATING SYSTEM: PC-DOS/MS-DOS
22 SOFTWARE: PatentIn Release #1.0, Version #1.25
23 CURRENT APPLICATION DATA:

```

```

1      APPLICATION NUMBER:  US/08/544,332
2      FILING DATE:
3      CLASSIFICATION:  435
4      PRIOR APPLICATION DATA:
5      APPLICATION NUMBER:  US 07/991,867
6      FILING DATE:  07-DEC-1992
7      PRIOR APPLICATION DATA:
8      APPLICATION NUMBER:  US 08/107,755
9      FILING DATE:  19-AUG-1993
10     PRIOR APPLICATION DATA:
11     APPLICATION NUMBER:  WO 92/14818
12     FILING DATE:  12-FEB-1992
13     PRIOR APPLICATION DATA:
14     APPLICATION NUMBER:  US 07/827,685
15     FILING DATE:  30-JAN-1992
16     PRIOR APPLICATION DATA:
17     APPLICATION NUMBER:  US 07/657,584
18     FILING DATE:  19-FEB-1991
19     ATTORNEY/AGENT INFORMATION:
20     NAME:  Bencen, Gerard H.
21     REGISTRATION NUMBER:  35,746
22     REFERENCE/DOCKET NUMBER:  UFI114,C4
23     TELECOMMUNICATION INFORMATION:
24     TELEPHONE:  904-375-8100
25     TELEFAX:  904-372-5600
26     INFORMATION FOR SEQ ID NO:  41:
27     SEQUENCE CHARACTERISTICS:
28     LENGTH:  1689 base pairs
29     TYPE:  nucleic acid
30     STRANDEDNESS:  single
31     TOPOLOGY:  linear
32     MOLECULE TYPE:  DNA (genomic)
33     US-08-544-332-41

```

```

Query March 2.9% Score 37.4; DB 2; Length 1689;
Best Local Similarity 46.3% Pred. No. 0.2; Mismatches 181; Indels 1; Gaps
Matches 157; Conservative 0;

Oy 946 ttaactgtttaagatgaacatcacacatatttctacagctttccaaagtccaacca 10050
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 113 tttactgtttttaccattatatttttccataattattatttcgtttaattcattatccta 172
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy 1006 atctaaagcttcaatcttgctcttttggaatgacgtcgtctttgctcggttttagact 10655
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 173 gttttttaaggtttgctagaatfatccaatattgaagaatatttaaaccaatttaaaatcacat 232
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy 1066 aaacacttaccactctgcacatgttgact-catgttcagctacaataacttccaaact 11242
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 233 ttttaaccttgatattcccttcaaaaatttaacaaagggttcaaatatttaattcttttcctctc 292
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy 1125 tgaacaaaatagttatttataatgtatatactctgcataactaaagaattgctttttccc 11844
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 293 gaatttttaaaattattttttacagattttctatatttttacacgaattatganaattattttta 352
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy 1185 cagcgctgttccatagtgactagcagatataaactggttataacctataacctatactcagtgta 12444
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 333 ttaattattttatattcagcagattggttaatttagtttttttttttttttttttttttttttt 412
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy 1245 actgcttaagtttttactgcagaaataataatactta 1283
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 413 ctttcataataaagatttcagagaaatttttttttttttttttttttttttttttttttttttt 451
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
us-07-991-867b-1
; Sequence 1, Application US/07991867B
; Patent No. 5476781
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: No. 5476781el Entomopoxvirus Expression System

```



---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: August 22, 2002, 02:09:50 ; Search time 1887.91 Seconds

(without alignments)  
9172.362 Million cell updates/sec

Title: us-09-758-017a-1

Perfect score: 1283

Sequence: 1 gacatccgcttgaataatg.....agaataataataatactta 1283

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:\*  
1: em\_estbta:\*  
2: em\_estbhm:\*  
3: em\_estlin:\*  
4: em\_estlmu:\*  
5: em\_estkov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlcc:\*  
9: gb\_estl1:\*  
10: gb\_estl2:\*  
11: gb\_hlcc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	381	29.7	696	9	AJ395572
2	340.2	26.5	1175	10	BM471137
3	317.8	24.8	822	10	BI823926
4	313.8	24.5	808	10	BE236591
5	309	24.1	869	10	BE105781
6	307.6	24.0	609	9	AU126319
7	298.4	23.0	881	10	BM449708
8	295.4	22.9	622	10	BI066406
9	293.4	22.9	622	9	AJ394089
10	284.6	22.2	548	9	AA290918
11	280.2	21.8	863	9	AL559968
12	274.4	21.4	788	10	BI651753
13	272.6	21.2	574	10	BI940552
14	267.2	20.8	768	10	BI657579
15	262.4	20.5	614	10	BE866388
16	260.8	20.3	1061	10	BM458045
17	259	20.2	731	9	AV705903

18	253	19.7	612	10	BI064298
19	250	19.5	672	10	BM104057
20	243.8	19.0	924	10	BE176725
21	243	18.9	874	10	BE099178
22	241	18.8	627	10	BE309210
23	233.2	18.2	835	9	AA068344
24	230.4	18.0	806	9	AL669156
25	229.6	17.9	555	10	H09366
26	223	17.4	407	9	AA401453
27	222.4	17.3	688	10	BI937645
28	217.6	17.0	484	10	BI064543
29	216.4	16.9	491	9	AA633967
30	214	16.7	531	9	AJ396948
31	213.4	16.6	915	10	BE793197
32	212.6	16.6	914	10	BE902908
33	209.8	16.4	539	10	BI979825
34	208.6	16.3	887	10	BM012071
35	207.4	16.2	654	10	BM103222
36	207.4	16.2	938	10	BE389571
37	205.6	16.0	498	9	AA068321
38	204.6	15.9	858	10	BE701762
39	202.4	15.8	629	10	BI691535
40	199.6	15.6	555	10	T78215
41	198.6	15.5	475	10	BE751706
42	197	15.4	621	10	BI980431
43	186.8	14.6	374	9	AA356048
44	186.6	14.5	403	9	AV706222
45	183.4	14.3	532	10	BI675591

## ALIGNMENTS

BI064298	BI064298
BM104057	fv42d11.x
BE176725	602313539
BE099178	601751033
BE309210	601097426
AA068344	mm53g01.r
AL669156	AL669156
H09366	y195a08..r1
AA401453	UI-HP-BKO
BI937645	de03e0f.x
BI064543	BI064543
AA633967	bl1a402.w
AJ396948	AJ396948
BE793197	601583610
BE902908	601676618
BI979825	ft76c08.y
BM012071	603636506
BM103222	fv42d11.y
BE389571	602414290
AA068321	mm53e01.r
BE701762	602129220
BI691535	603307201
T78215	yc98g05..r1
BE751706	203907.MA
BI980431	ft76c08.x
AA356048	EST64540
AV706222	AV706222
BI675591	dac74d03.

RESULT 1  
AJ395572 696 bp mRNA linear EST 25-JAN-2001  
DEFINITION AJ395572 dkfz426 Gallus gallus cDNA clone 22p14r1, mRNA sequence.  
ACCESSION AJ395572  
VERSION AJ395572.1 GI:7126625  
KEYWORDS EST.  
SOURCE chicken.  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
REFERENCE 1 (bases 1 to 696)  
Abdulkhanov, I., Lodygin, D., Gerolt, P., Arakawa, H., Law, A., Plachy  
AUTHORS 'J., Korn, B. and Buerstedde, J.M.  
TITLE A large database of chicken bursal ESTs as a resource for the  
analysis of vertebrate gene function  
JOURNAL Genome Res. 10 (12), 2062-2069 (2000)  
MEDLINE 20568495  
COMMENT Contact: Buerstedde JM  
Cellular Immunology  
Heinrich-Pette-Institute  
Martinstr. 52, 20251 Hamburg, Germany  
Email: <http://genetics.hpi.uni-hamburg.de/dt40est.html>.

FEATURES  
source location/Qualifiers

1. 696  
/organism="Gallus gallus"  
/strain="CB"  
/db\_xref="taxon:9031"  
/clone="22p14r1"  
/clone\_lib="dkfz426"  
/tissue-type="Bursa of Fabricius"  
BASE COUNT 156 a 187 c 205 g 146 t 2 others  
ORIGIN

Query Match 29.7%; Score 381; DB 9; Length 696;  
Best Local Similarity 72.9%; Pred. No. 3.7e-83;  
Matches 489; Conservative 0; Mismatches 182; Indels 0; Gaps 0;

OY	252	ccctgagagcttgcggagagccttcggagagaagagcgtgcgcgcgaggtttgaaagacatc	311
Db	26	ccctccggcgtttcgggagagcgttcggcggccgacagcttgcggcagagttcaccaagccctac	85
OY	312	ttcaacaacattgatctcctttgtaqctga tggagagagcgcgtcacacacgcttaccacacg	371
Db	86	tttcgtagacattgatggcatttcgtggccgagagaggaagaggtatgacgcggtatcccgccc	145
OY	372	gctgattcaagtttacaagtttcgcacagatagtttgacattcaagaatgtgaaagtatgat	431
Db	146	ccggagcagagcttctcacttggacacagatgtcgcacatcaggaatgaggaattgtgtaatt	205
OY	432	ctagagccaaagaccccttaccacagctcccaacaaagacacatcgaactcgttccagttgcga	491
Db	206	ctgggacaaagattccttatgatgatgacctatcagaagctcatgggctctgcttccagtgcacg	265
OY	492	aagcgaattcccccctcccccagctcgtgtgaacatatataaagaattgtgtccgacatt	551
Db	266	aaacctgtgtccacctcccccctccagtttgaaaaacattttacaaaagaaactgtcagatgatt	325
OY	552	gatgcttcaagcatcccttgacatctgagatctcaagcggatgggacaaagaggggtgcg	611
Db	326	gaagacattcacccattcttggtacggcattctgactgcttgaggccacagcagaggtgtcttc	385
OY	612	ctgcttaacgcggtgtcgtgacgctgcgggcccacacagccaactcccaagaagacagagc	671
Db	386	ctgctcaacgcgcgcgtcctcaccggtggagccacacagccacgcctccacacagagagggc	445
OY	672	tggagagaccttaccgcagcgttgatcaagttgactgagtcgagtcacacgggaaggaatcgt	731
Db	446	tggagagcagattcacggagtgtgctcgtcttcttggtgacaaacaaacctgcattggggtgtt	505
OY	732	ttctctgtgtggtgctcatabgcaccataagaaggagcgaccatcgacagaaagcagtcac	791
Db	506	tttcacgctctggggagcgttatcacaagaagggcagcagcttattgacaggaagccac	565
OY	792	catgctctgcaagcgtgtcatcactcctcttgtgtgcacatcgtggtgttccttgatgt	851
Db	566	cacgcttctgcacacagcgttcccttcccttctcttctgtgaaacagagggctcttggcgtt	625
OY	852	aagcactcttccaaagcttaacgcggtgtcgttgaactatcttggagcagagcctataacgtg	911
Db	626	cggcacttcttctcacaagaacagagcgtgtaagaagtgctgcgcnaagagccatcgacttg	685
OY	912	agagcactccta 922	
Db	686	agagcgtgcta 696	
RESULT	2		
LOCUS	BM471137	1175 bp	mRNA linear EST 05-FEB-2002
DEFINITION	AGENCOURT.6463958 NIH_MGC_71	human	CDNA clone IMAGE:5517670
ACCESSION	BM471137		
VERSION	BM471137.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 1175)		
JOURNAL	NIH-MGC http://mhc.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cga@bms.femail.nih.gov		
	Tissue Procurement: ATCC		
	CDNA Library Preparation: Life Technologies, Inc.		
	DNA Sequencing by: Agencourt Bioscience Corporation		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNLN at:		

[illegible]



Db	332	CGGTTCCGTCGCCATTCACACCAACTCTCATTAAGAGACCGAGCTGGAGACAGTTTCACTATG	391
QY	691	ctgtgataaagtgctgcgcgtccaaaccgggaagagtcgtcttccctgtgtgaggctcat	750
Db	392	CAGTTGTCTCCTGGCTAATACAGACTCGAATGGCCCTTGTTTCTTGGCTCTGGGCTCTT	451
QY	751	acgccataaagaaggagcgagaccatcgacaggaaacgctacacatgtcttgcgaagctgttc	810
Db	452	ATGCACAAAGAAAGGCGAGTGCATTGATAGAACCGGACCATCTACTACGACGAGGCTC	511
QY	811	atccatctcccttgcgtcctcatcgttgaggttcccttgcgtgtgaagactctcccaagcta	870
Db	512	ATCCCTCCCTCTTGTGTCACTGTTATAGAGGCTTCTTTGGATGTAAGACACTTTTAAAGACA	571
QY	871	acggagctgcgtcaaaactatcctggagcagcgactataaactgagagagcaactctaa	923
Db	572	ATGAGCTCTCTGCAGAGTCTGGCAAGAACGCCATTGACTGGAAGAGAGCTGTGA	624
RESULT	4		
LOCUS	BF236591	808 bp	mRNA linear EST 14-NOV-2000
DEFINITION	60202877.F1 NCL_CGAP_L19	Mus musculus	CDNA clone IMAGE:4163920 5'
ACCESSION	BF236591		
VERSION	BF236591.1	GI:11150385	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus. 1 (bases 1 to 808)		
AUTHORS	NIH-MGC <a href="http://mgc.ncl.nih.gov/">http://mgc.ncl.nih.gov/</a> .		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgabs@mail.nih.gov">cgabs@mail.nih.gov</a> Tissue Procurement: Jeffrey E. Green, M.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMNL at: <a href="http://image.lnl.gov">http://image.lnl.gov</a> Plate: LLM9448 row: n column: 17 High quality sequence stop: 708.		
FEATURES	Location/Qualifiers		
SOURCE	1..808		
	/organism="Mus musculus"		
	/strain="FVB/N"		
	/db_xref="taxon:10090"		
	/clone_1b="IMAGE:4163920"		
	/clone_1lb="NCL_CGAP_L19"		
	/lab_host="DH10B (T1 phage-resistant)"		
	/note="Organ: liver; Vector: PCMV-SpOR6; site_1: NotI; site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.9 kb. Constructed by Life Technologies. Note: this is a NCL_CGAP Library."		
BASE COUNT	191 a	217 c	207 g 193 t
ORIGIN			
Query Match	24.5%;	Score 313.8;	DB 10; Length 808;
Best Local Similarity	69.5%;	Pred. No. 1,4e-66;	
Matches 426;	Conservative 0;	Mismatches 187;	Indels 0; Gaps 0;
QY	312	ttcaacaatgatgtccttctgtagctgaatgagagcgctcaacacgctctaccacg	371
Db	111	TTAAACCGAGTAAATGGGATTGTGTGCTGAAGAAAGAACACCAACCAAGCTCATCCGCC	170
QY	372	gctgataaagtgatacgtctgcacagagatgctgatacatcctaagtgtgaagtgtgatt	431
Db	171	CCGAGGACGGGTTCACATGGAGCCAGATGGCGACATCCAGATGTGAAGTGTGTCTATT	230

Db	Accession	Source	Organism	Reference	Authors	Title	Journal	Comment
QY	432	ctaggcgaagacccttaccacggtcccaacacgaacatggaactgtgttcagtgtgca	869 bp	EST	30-JAN-2001			
Db	231	ctgggacgagatccctatcattacgagacttattcaagctcagggctctgcttccagtgtccaa	290					
QY	492	aagcaggttcccccctcccccagctcgtgaaacatacaaaaagtgtgtacacgacatt	551					
Db	291	aaacacagttccacccccccgcttggcaaaacatttttaaaagagctgctcacacatc	350					
QY	552	gattgcttcaaacatctctggaatgagatctaaagcagatggtgcaaaaaggggtgtcgt	611					
Db	351	gattggtttgtttcatctctggtccatggtgggattgtgtcaggggtggcccgacaaaggtgtcttc	410					
QY	612	ctgcttaacgcyggtgtcgcacggtgcgggcccatacgaagccaaactcccaacaggacagagtc	671					
Db	411	cttcctcaacgacgctgctctcactgttcggccgccacaaagccaatgtcccatgaagagagggcc	470					
QY	672	tgggagaccttaacgagacgtgtgatacaagtgtgcgaagctcaaccgggaagagatcgtt	731					
Db	471	tgggagacattacgagacgctggtgtgtctgctgctgaatcagaacctgagtgccctcgtc	530					
QY	732	ttccggtgtggtgggtctcaacacgcccataaaggagagcagcatcgacaggaacgtcac	791					
Db	531	tttccttccttggggcgttattgctcagagaagggcagtgctattgacaggaagcctcac	590					
QY	792	catgtctcgcgaagctgttcatccatctcttgtctgtctcagtcgtgtgttctcgtgtgt	851					
Db	591	catgttctgcacacagcttccacccctccgctgcgtgtacagaggggttcttgggattcc	650					
QY	852	aagcactctccaaagctaacgagctgcgtcgtgaactctctggagcagagccttaactgtg	911					
Db	651	aaacatttttttcaaaagccattgaactgcttccagaaagcttggcagaagcccatcaactgg	710					
QY	912	agagacacttaac	924					
Db	711	aaggaaggtgtgac	723					
RESULT	5							
LOCUS	BG105781	869 bp	mRNA	linear	EST	30-JAN-2001		
DEFINITION	602312622P1 NIH_MGC_84 Homo sapiens CDNA clone IMAGE:4422292 5',							
ACCESSION	BG105781							
VERSION	BG105781.1	GI:12599627						
KEYWORDS	EST.							
SOURCE	human.							
ORGANISM	Homo sapiens							
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.							
TITLE	1 (bases 1 to 869)							
JOURNAL	NIH-MGC http://mgs.nci.nih.gov/							
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)							
	Unpublished (1999)							
	Contact: Robert Strausberg, Ph.D.							
	Email: cga@bbs.femail.nih.gov							
	Tissue procurement: ATCC							
	cDNA Library Preparation: Life Technologies, Inc.							
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)							
	DNA Sequencing by: Incyte Genomics, Inc.							
	Clone distribution: MGC clone distribution information can be							
	found through the I.M.A.G.E. Consortium/LNL at:							
	http://image.lnl.gov							
	Plate: L14M10162 row: 1 column: 05							
	High quality sequence stop: 653.							
FEATURES	Location/Qualifiers							
source	1..869							
	/organism="Homo sapiens"							
	/db_xref="taxon:9606"							
	/clone="IMAGE:4422292"							
	/clone_lib="NIH_MGC_84"							
	/tissue_type="adrenal cortex carcinoma, cell line"							
	/lab_host="DH10B (phage-resistant)"							

BASE COUNT 221 a 200 c 225 g 223 t  
ORIGIN

/note="Organ: adrenal gland; Vector: PCMW-SPO6; Site\_1:  
NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dr  
primed. Average insert size 1,229 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH-MGC library."

[illegible]

Genomics Laboratory  
 Helix Research Institute  
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
 Tel: 81-438-52-3951  
 Fax: 81-438-52-3952  
 Email: genomics@hri.co.jp  
 HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix  
 Research Institute; cDNA library construction: Department of  
 Virology, Institute of Medical Science, University of Tokyo, and  
 Helix Research Institute.

# FEATURES

source  
 Location/Qualifiers  
 1. 609  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="NT2RP1000586"  
 /clone\_1ib="NT2RP1"  
 /cell\_type="teratocarcinoma"  
 /cell\_line="NT2"  
 /note="Vector: pUC19FL3; mRNA from NT2 neuronal precursor  
 cells after 48-hours retinoic acid (RA) induction"  
 BASE COUNT 152 a 142 c 144 g 163 t 8 others  
 ORIGIN

Query Match 24.0%; Score 307.6; DB 9; Length 609;  
 Best Local Similarity 69.8%; Pred. No. 4.3e-65;  
 Matches 423; Conservative 0; Mismatches 182; Indels 1; Gaps 1;

310 acttcaacaattgattcttcttgtagctatgagagagcgccacacgcttaccac 369  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 4 attttatcannctnattgattgattgttcacagaaagaaagcattacacgtttatccac 63  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 370 cggctgacagtgatgacagtcagacagagatgltgacatcacaagtgtgaaagtagtga 429  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 64 cccacacacagatcttccac 123  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 430 ttctagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 489  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 124 tcttgaggag 183  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 490 aaaaagcagctcccccac 549  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 184 aaagcctgttcccccac 243  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 550 ttgatgctcacaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 609  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 244 tagagattttgttcatcctgagcagcagcagcagcagcagcagcagcagcagcagcagcag 303  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 610 tgcctgctaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 669  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 304 tcccttctcaacgctgcttccac 363  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 670 gctgagac 729  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 364 gcggagac 423  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 730 ttctcctgtgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 789  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 424 ttttcttgcctgctggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 483  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 790 acatgctgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 849  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 484 accatgctgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 543  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 850 gtaagcactctcacaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 909  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 544 gttnnacccttttcaaaac 602  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 910 gggagag 915  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 603 ggaang 608

RESULT 7  
 LOCUS BM449708 881 bp RNA linear EST 05-FEB-2002  
 DEFINITION AGENCOURT\_6392952 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:5528414  
 5', mRNA sequence.  
 ACCESSION BM449708  
 VERSION BM449708.1 GI:18498748  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS NIH-MGC  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabs@femail.nih.gov  
 Tissue Procurement: ATCC/DC/DMP  
 cDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LLM12205 row: d column: 15  
 High quality sequence stop: 623.

# FEATURES

source  
 Location/Qualifiers  
 1. 881  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5528414"  
 /clone\_1ib="NIH\_MGC\_72"  
 /tissue\_type="melanotic melanoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: skin; Vector: pCMV-Sport6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 2 kb. Library constructed by Life  
 Technologies."  
 BASE COUNT 219 a 209 c 220 g 231 t 2 others  
 ORIGIN

Query Match 23.3%; Score 298.4; DB 10; Length 881;  
 Best Local Similarity 71.0%; Pred. No. 8.6e-63;  
 Matches 395; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

368 accggtgatcaagtgatgacagtgatgacagagatggtgacattcaaatgtgaaagtgtg 427  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 4 acgggtggcccaagcttccac 63  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 428 gatctagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 487  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 64 catcttgagacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 123  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 488 gcaaaagcagcttcccccac 547  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 124 tcaaaagcagcttcccccac 183  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 548 catgatgctcacaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 607  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 184 catgagcag 243  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 608 gctgcttcaacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 667  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 244 tctcttctcaacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 303  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 668 agcctgagcag 727  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 304 agcctgagcag 363  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 728 cgttctcgt 787  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB	364	TTTTTTCTTGCTCTGAGGCTCTTTATGCTCAGAGAAGGGCACTGCCATTGATGAGAACGC	423
OY	788	tcaccatgctctgaagaagctgttcacatccatctctgtctgcatacgttgccttg	847
DB	424	GCACCATGACTACAGAGAGGCTCATCCCTTTCCTTTGTCAGTGTAFAGAGGGTCTTTGG	483
OY	848	ttgtaagcactctccaagagctcaaggctctgtcgttaactatacttggagcgagctataa	907
DB	484	ATGTGAGACACTTTTAAAGAACCAATGAGCTGCTGCAGAAAGTCTGGCAGAAAGCCCATTTGA	543
OY	908	ctggagagcaactctaa	923
DB	544	CTGGAAGGAGCTGTGA	559
RESULT	8		
LOCUS	B1066406	622 bp	mRNA
DEFINITION	pgfln.pk008.a20	normalized chicken fat cDNA library	EST 15-JUN-2001
ACCESSION	B1066406		
VERSION	B1066406.1	GI:14473928	
KEYWORDS	EST.		
SOURCE	chicken.		
ORGANISM	Gallus gallus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.		
AUTHORS	I (bases 1 to 622)		
TITLE	Cogburn, L.A., Morgan, R.W. and Burnside, J.		
JOURNAL	Chicken ESTs from fat		
COMMENT	Unpublished (2001)		
FEATURES	Contact: Larry A. Cogburn		
SOURCE	University of Delaware		
	Townsend Hall, Newark, DE 19717, USA		
	Tel: 302-831-1335		
	Fax: 302-831-2822		
	Email: cogburn@udel.edu, www.chickest.udel.edu.		
	Location/Qualifiers		
	1..622		
	/organism="Gallus gallus"		
	/db_xref="taxon:9031"		
	/clone="pgfln.pk008.a20"		
	/clone_lib="normalized chicken fat cDNA library"		
	/sex="Male and female"		
	/tissue_type="fat"		
	/lab_host="E. coli EMDH10B"		
	/note="Vector: pSPORT1"		
BASE COUNT	137 a	165 c	194 g
ORIGIN		119 t	7 others
Query Match	23.0%	Score 295.4;	DB 10; Length 622;
Best Local Similarity	70.0%	Pred. No. 4.3e-62;	
Matches 392; Conservative	0;	Mismatches 168; Indels	0; Gaps 0;
OY	186	gagcagctgggaagaatgcccacaaataagaagaagcagcgtcttgacaagaatagagcaaa	245
DB	63	GAGCAGCTGGAGAGCGGATCGCAGGAACAAAGAGGTAGCGCTGCGGAGGCTCAACGAGCGC	122
OY	246	gcaaacgctgcaggttgcgagagactgcgaggaagagagcgtgcgcagagtttgaagaag	305
DB	123	AGCGTCTCCGGGCTTGGGGAGAGCGTGGAGCGGCCACGCTGGCGCCAGATTACCAAG	182
OY	306	ccatacttcaacaatgatgatgttcctttgtagctgatgagagagccggtcacacagctctac	365
DB	183	CCCTACTCTCGTAGAGCTGATGGCATTCGTGGCCGAGGAGAGAGAGGTACACGCGTTCAC	242
OY	366	ccaccgctgatacagtgatacagtttcgacagagatggtgacatccaagaatgtaagaatga	425
DB	243	CGCGCCCGGAGGAGGCTTCACTTGGACACAGATGTCCGACATCAGGAGGATGTAAGATTT	302

OY	426	gtgatctctatgagccaggacccttaacacaggttcccacccaagaacacatggaactcgttttcagt	485
Db	303	GTAATTCCTGGGACCAAAATCCTTATTATCATGTAGACCCTAAATCAAGACTCAGTGCTCTGTTTCACT	362
OY	486	gtgcacaagaacaggttccccctcccccaggttcctcgtagacatatcaaaagaattgtgtacc	545
Db	363	GTCCAGAAACCTGTCTCCGCTCTCCCCCAATTGGAAAAAATATTACAAAACACTGTGTGACT	422
OY	546	gacattcatgctctcaaatccttgagacatlyggagatactaagcggatgggcaaaaagaag	605
Db	423	GATATGGAAGACTTCACCATCTCGGTATCGGTATCGAGGATCTAGTCAGGCTGGGCCAACGAGGCT	482
OY	606	gtgtcgtctcttaaagcaggttgtctgacagctgtggcgggcccatcaggtcccaactccacaagaagc	665
Db	483	gtgtctctctctCAAGCCCTCTCTCAAGGTGAGGCCACAGCCACGTCACCAAGAAGAG	542
OY	666	agagagcttgtagaacctcacacagcgtgtgatacagttgacctgagctgaacccgggaaga	725
Db	543	AAGGCTGGGAGACAGTTCAACGATGTGTGCTCTCTGCTGCTAACAAGAACCTGCNGNNN	602
OY	726	gtcgtttccctcgtltgtl99g 745	
Db	603	NNMTGTTTCATGCTGTMGGG 622	
RESULT	9		
LOCUS	AJ394089	792 bp	mRNA linear EST 25-JAN-2001
DEFINITION	AJ394089 dkfz426 Gallus gallus cdna clone l7p6rl, mRNA sequence.		
ACCESSION	AJ394089		
VERSION	AJ394089.1 GI:7123633		
KEYWORDS	EST.		
SOURCE	chicken.		
ORGANISM	Gallus gallus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Archosauromia; Aves; Neognathae; Galliformes; Phasianidae;		
	Phasianinae; Gallus.		
REFERENCE	1 (bases 1 to 792)		
AUTHORS	Abdrikmanov,I., Lodygin,D., Gerolt,P., Arakawa,H., Law,A., Plachy		
TITLE	J., Korn,B. and Buerstedde,J.M. A large database of chicken burral ESTs as a resource for the analysis of vertebrate gene function		
JOURNAL	Genome Res. 10 (12), 2062-2069 (2000)		
MEDLINE	20568495		
COMMENT	Contact: Buerstedde JM Cellular Immunology Heinrich-pette-Institute Martinistr. 52, 20251 Hamburg, Germany Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.		
FEATURES	Location/Qualifiers		
Source	1..792		
	/organism="Gallus gallus"		
	/strain="CB"		
	/db_xref="taxon:9031"		
	/clone="l7p6rl"		
	/clone_lib="dkfz426"		
	/tissue_type="Bursa of Fabricius"		
BASE COUNT	170 a 225 c 247 g 148 t		2 others
ORIGIN			
Query Match	22.9%	Score 293.4:	DB 9; Length 792;
Best Local Similarity	70.1%;	Pred. No. 1.4e-61;	
Matches 407; Conservative	0;	Mismatches 173;	Indels 1; Gaps 1;
OY	186	gagcagctggaagaatgagccaaaaataagaagaagcagcttgacaagattagagcaaaa	245
Db	206	GAGCAGCTGGAGCGGATGCGCAGGAAACAAGAGTAGCGCTCCGAGGCTGACAGAGCGC	265
OY	246	gcaacgcctgcaggttctcgagaagacttgagagaagagagctgtgctgcagagtttgaag	305
Db	266	AGCGTCTCTCCGGGCTTGGGGAGAGCTGGCGCGCGAGCTGGCGCGAGATTCCACGAAG	325



FEATURES	SOURCE
Location/Qualifiers	
1. 863	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone="CSODG004Y119"	
/clone_11b="L11_FL011_BC1"	
/sex="male"	
/tissue="type="B" cells from Burkitt lymphoma"	
/lab_host="DH10B"	
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : <a href="mailto:liang@lifetech.com">liang@lifetech.com</a> URL : <a href="http://fulllength.invitrogen.com">http://fulllength.invitrogen.com</a> "	
BASE COUNT	188 a 242 c 250 g 182 t 1 others
ORIGIN	

Query Match	21.8%	Score 280.2	DB: 9	Length 863
Best Local Similarity	66.2%	Prod. No. 2.6e-58		
Matches 419	Conservative 0	PredMatches 213	Indels 1	Gaps
OY 159	aagaacatcgtcccaacgaatttcagttgtagcgagctcggaagaagatgcccnaataaga	218		
Db 218	ACGCCGCCCTCTCCCGTCGTAGTGCACAGCATGTTGGACCGGATCCAGAGAAACAAGCC	277		
OY 219	gcagcgcttgcgaagaattagagcaaaagcaacgcctgcagtttcggaagacgttgaga	278		
Db 278	CGCGCCCTGCTCAGACTCGCGCCGCCGCAACGTCGCCCGTGGCTTTGGAGAGAGCTGGAAG	337		
OY 279	agagcgctgcgcgcagagtttggaaaagcgcatacttcaacaatltgatgtccctttagct	338		
Db 338	AAGCACCCTCAGCGGGGAGTTCGGGAAACCGATTTTATATCAACCTATAGGATTTGTGGA	397		
OY 339	gatgagagagcgctgcacacccgtctaccacccgcgctgatacagttgtacagttcgacag	398		
Db 398	GAAAGAAAGAACATTTACACTGTTATTCACCCCCACACCAAGTCTTCACTCGAGACCGAG	457		
OY 399	atgtgtgacatcaagaatgtgaaglaatgaltcttaagccagagacccttaacaggtccc	458		
Db 458	ATGTGTGACATTTAAAGATGTGAAGGTTTGATCTCGGGACAGATTCATATCATGTGACCT	517		
OY 459	aaccagacacatgagctctgttttagttgttgcaaaagcgaattcccccctcccgagctcc	518		
Db 518	AATCAAGCTCACGGGCTCTGCTTATGATGTTCAAAGCCCTGTTCGCCCTCCGCCAGTTTGG	577		
OY 519	gtgaacatacaaaagaattgttaccgcacatgtgatggtcttaagaacatccctgaacatga	578		
Db 578	GAGAACATTTTAAAGAGCTGTGCTACAGACATAGAGGATTTGTTCATCTCTGGCCATGGA	637		
OY 579	gatctaagcgaatggtggcaaaagaagggctgctgtcgttaacgcggtgtgtgacggctg	638		
Db 638	GATTTATCTGGGTGGGGCAAGCAAGTGTTCCTTCATCAACGCTGCTCCTCACGGTTCGT	697		
OY 639	gcccatcagcgccaactcccaacaagaacagagctgtggagaccttaccagacgtgtgac	698		
Db 698	GCCCATCAAGCCACACTCTCATTAAGGAGGCGAGGCTTG6GACCACTGATCACTGATTGG	757		
OY 699	aagtgtgcgaagcgcttaaccgggaagagtgcttttcctgttctgtggggtcctataagcccat	758		
Db 758	TCTGTGGCTTAATCACAAACTCCGAATGGCCTTGTCTTCTTGCTGTGGGGCTCTTATGCTCAG	817		
OY 759	aagaagagagcgacacatgcagcagaagaacgtcac	791		
Db 818	AAGAAGGG-CAGTCGATTGATAGGAAGGGCGAC	849		

LOCUS	B1651753	788 bp	mRNA	linear	EST_12-SEP-2001
DEFINITION	6032965541F1 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5339027 5',				
ACCESSION	mRNA sequence.				
VERSION	B1651753				
KEYWORDS	B1651753.1	GI:15565989			
SOURCE	EST.				
ORGANISM	house mouse.				
TITLE	Mus musculus				
REFERENCE	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Eumetazoa; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
JOURNAL	1 (bases 1 to 788)				
COMMENT	NIH-MGC http://mgc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) unpublished (1999)				
	Contact: Robert Strausberg, Ph.D. Email: cgapbsr@mail.nih.gov Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN). DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.lnl.gov Plate: LHAM1861 row: a column: 12 High quality sequence stop: 788.				
FEATURES	Location/Qualifiers				
source	1..788				

FEATURES		Location/Qualifiers	
Source			
	1. .788		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/clone="IMAGE:5339027"		
	/clone_lib="NIH CGAP_Mam3"		
	/tissue_type="tumor, gross tissue"		
	/lab_host="DH10B"		
	/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1, Nct1; Site:2: Salt; Cloned unidirectionally. Primer: Oligo dT. Average insert 2 kb. Library constructed by Life Technologies, catalog #12017-018. Investigators providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference for transgenic model: Xu et al., Nature Genetics 22, 37-43 (1999). Note: this is a NCI_CGAP Library."		
BASE COUNT	174 a 236 c 206 g 172 t		
ORIGIN			
Query Match	21.4%;	Score 274.4;	DB 10%; Length 788;
Best Local Similarity	70.2%;	Pred. No. 6.4e-57;	
Matches 368;	Conservative 0;	Mismatches 156;	Indels 0; Gaps 0;
QY	403	gtgacatccaagatgtgaaagtatgattcttagcgacgaagaccttaccacggtcccaacc	462
Db	13	ggcgaatccgagatgtgaaggtttgattcttcggacagagatcccatTAACGACCTAATC	72
QY	463	aagcacatgagactgtttcagltgacaaagcaagttccctcccccagttctcgtga	522
Db	73	AAGCTCAGGGGCTGCTTCAGTGTCCAAAGACAGTTCACACCCCGCCAGTTTGGAAA	132
QY	523	acatatacaagaatgtgtacccagcatgtatgagttcaagaatctctgagatgagatc	582
Db	133	ACATTTTAAAGAGCTGTCTACAGCATTCAGATGGTTTGTTCATCTGGCCATGGGATTT	192
QY	583	taagcgaatgggcaaaaaggggtgcgtctgcttaacgcggtgtctgacgtgcggccc	642
Db	193	TGTTCAGGGTGGGCCCGACAAGAGTGTCTCTCTCAACGCGCTCTCACTGTCCGCGCC	252
QY	643	atcagcgaactcccaacagacagagcctggagagaccttaccogaagcgttgatcaagt	702
Db	253	ACCAAGCCATTTCCCATTAAGAGAGAGGGCTGGAGACAGTTCCAGGACCGGGCGTGTCT	312
QY	703	ggctgagcgctcaacgcgggaagagtcgtttctctgtgtggtggtctaatggccataaga	762
Db	313	GGCTGAATCAAGACTGATGGCTCTGTCTTCTCTCTGGGGCTTTATGCTCAGAAAGA	372
QY	763	aggagagcaccatcgacaggaagacgttcaacatgtcttgcaagcgtttatcatcatctct	822

|||||  
Db 373 AGGCGATGTCATTTGACAGAGAGCGTCACCATGTTCTGCGACAGACGTCACCCCTCCCGC 432  
QY 823 tgcgtcgcacgctgggtctcgttctgttaagaagctcccaaggtcaagcgtgctgta 882  
Db 433 TGTGCGGTGTCACAGAGGCTTCTCGGATGACAGACATTTTCTAAAGCCATGAACTGCTCC 492  
QY 883 aactatctggagcggagcctataactgagagacactcacc 926  
Db 493 AGAAGTCTGGCAAGAACCCCATCACTGAAAGAGCTGTGACCC 536

## RESULT 13

LOCUS B1940552 574 bp mRNA linear EST 18-OCT-2001  
DEFINITION dbj1106.y1 Wellcome CRC PSK egg Xenopus laevis cDNA clone  
IMAGE:3378467.5 similar to SW:UNG\_HUMAN P13051 URACIL-DNA  
GLYCOSYLASE PRECURSOR ; mRNA sequence.

ACCESSION B1940552  
VERSION B1940552.1 GI:16255024

KEYWORDS EST.  
SOURCE African clawed frog.  
ORGANISM Xenopus laevis

REFERENCE 1 (bases 1 to 574)  
AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,  
Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person  
B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,  
Waterson,R. and Wilson,R.  
Washu Xenopus EST project, 1999  
Unpublished (1999)

TITLE Amphibia: Batrachia: Anura: Mesobatrachia: Pipidae: Pipidae;  
JOURNAL Xenopodinae: Xenopus.  
COMMENT Other\_ESTS: dbj1106.x1

Contact: Sandy Clifton, Ph.D.  
Washu Xenopus EST project, 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Library constructed by N. Garrett, P. Lemaire, A.M. Zorn, and J.B.  
Gurdon (Wellcome/CRC Institute). DNA sequencing by: Washington  
University Genome Sequencing Center  
Clone distribution: Xenopus clones from this library are available  
through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov  
Seq primer: -40RP from Gibco  
High quality sequence stop: 454.

## FEATURES

## source

1..574  
/organism="Xenopus laevis"  
/db\_xref="taxon:8355"  
/clone="IMAGE:3378467"  
/clone\_lib="Wellcome CRC PSK egg"  
/tissue\_type="egg"  
/lab\_host="DH10B (phage-resistant)"  
/note="Vector: pBluescript SK-; Site\_1: NotI; Site\_2:  
EcoRI; CDNAS were oligo-dn primed and directionally  
cloned. Library was constructed by N. Garrett, P. Lemaire,  
A.M. Zorn, and J.B. Gurdon (Wellcome/CRC Institute).  
Note: This is a Xenopus Gene Collection (XGC) library."  
BASE COUNT 170 a 132 c 136 g 136 t  
ORIGIN

Query Match 21.2%; Score 272.6; DB 10; Length 574;

Best local Similarity 67.5%; Pred. No. 1.8e-56;

Matches 383; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

QY 273 tggagaagagagctgctgcagagtttgaagccactactcaaacatgatgctcctt 332  
|||||  
Db 8 TGAAGACAGAGAGCTGCTGAGTTGCAAAACCTACTTTTGCAAGCTATCAAAATTTTC 67

QY 333 gtatgtatgagaggcgctcaaccgctctaccaccggctgataagtgtacgttcg 392  
|||||  
Db 68 ATTGCAGAAAGAGAAAGAAATGCGACAGTTTATTCACACACAGAGGATTCACCTGG 127  
QY 393 acagagatgtgtacatccaagatgtgaagtagtgaattcagaaccagaccctaccac 452  
|||||  
Db 128 ACACAAATGTTGATCACAAGATGTAAGAGTTGTTATTTTAGGCCAAGATCCTTATCAT 167  
QY 453 gttcccaaccagacatgacatgtcttccagltgtcaaaaagccagttccctccccc 512  
|||||  
Db 188 GGAACCAATCAAGCTCATGTGCTTTGCTTAGTGTCAAGAAAGCAGACCTCCACCC 247  
QY 513 agtctgtgaacatatacaagaattgtacacgacttgatgcttcaagatccgtga 572  
|||||  
Db 248 AGCTTAGTAAATATGTAAGAAAGTAAGAACCCATATCGAAGGATTTTCACGCCAGGT 307  
QY 573 catgaaatcaagcagatgtggcaaaaggggtgtctgtcttaacgggtgtgac 632  
|||||  
Db 308 CATGGGATCTCACAGGCTGGGCCAAACAAAGAGATCCTCCTCATATGCTGTGACA 367  
QY 633 gtgcgggccatcaagcgaactcccaagaagagagcttggagaccctcaaccgact 692  
|||||  
Db 368 GTGCGAGCTCATATATGCAAACTCTCACAGGACTGTGATGGGAACAGTTCCACAGATGT 427  
QY 693 gtgataagtggtgctgagcgtcaaccgggaaggagtgctgtgtgtgtgtgtac 752  
|||||  
Db 428 GTGGTATCTGGCTAAATTAAGACATGATGGTTGTTTCATGCTGTGGGAGCCTAT 487  
QY 753 gccataaagagagagacatcagacaggaagacgtacatgcttgcgaagcgtgcat 812  
|||||  
Db 488 GCCCAGAAAGAGCAGACCATATTTGACAGAAAGGACCATGTTCTTCAAAACCGTGCAC 547  
QY 813 ccatctcctgtctgtcgtcgtggtg 839  
|||||  
Db 548 CCTTCTCTCTCTCATGTCACAGAGGG 574

## RESULT 14

## LOCUS

B1657579 768 bp mRNA linear EST 12-SEP-2001  
DEFINITION 603281538F1 NIH\_CGAP\_Mam4 Mus musculus cDNA clone IMAGE:5325848.5',  
mRNA sequence.

ACCESSION B1657579  
VERSION B1657579.1 GI:15571815

KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 768)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: Lotmar Hennighausen Ph.D., Priscilla Furth  
Ph.D.

CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LRAM1826 row: 1 column: 09  
High quality sequence stop: 764.

## FEATURES

## source

1..768  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5325848"  
/clone\_lib="NIH\_CGAP Mam4"  
/tissue\_type="tumor, gross tissue"  
/lab\_host="DH10B"



/note="Organ: mammary; Vector: PCMV-SPORT6; site\_1: NotI; site\_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert 2.5 kb. Library constructed by Life Technologies, catalog # 12018-016. Investigators providing samples: Lothar Hennighausen/Priscilla Furth, NIH Reference for transgenic model: Li et al., Cell Growth and Differentiation 7, 3-11 (1996). Note: this is a NCI\_CGAP Library."

BASE COUNT 148 a 236 c 238 g 146 t  
ORIGIN

Query Match 20.8%; Score 267.2; DB 10; Length 768;  
Best Local Similarity 66.7%; Pred. No. 4.1e-55;  
Matches 411; Conservative 0; Mismatches 203; Indels 2; Gaps 2;

QY 150 gtggaacaaagacgcatcgccacagcttcctagtgagagctggaagaatgagccaa 209  
DB 154 GTGGAACAAAGACGACAGGAGGCTCGCGCTCAGCGCCGAGAGCTCGTCCGATCCAGAGG 213  
QY 210 aataaagaacagcgtctggaagaattagagcaaaagcaacgctgcaggttcgagag 269  
DB 214 AACAAAGGCGCGGCGCTGCTCAGAGCTCGCCCGCAAGCTGCTGCGGCTTCGGGAG 273  
QY 270 acttgagaagagagctgctgacaggtttgaaagccacttcaacaattgagtcc 329  
DB 274 AGCTGGAGCAGCAGCTGTGTCGGGGAGTTGGGAGCCGCTACTCTGCTCAAGCTAATGGA 333  
QY 330 ttgtgagctgagagagagcgctcaacacgcttaccacagcgctgagcaagtgaag 389  
DB 334 TTGTGCTGCTGAAGAAGAACACACACAGCTATCTCCGCCCGGAGAGGTTCACA 393  
QY 390 tcgacagaagatgtgtacatcacaagatgtgaagtagtgaattcaggccagaccttac 449  
DB 394 TGACCCAGATGTGCGACATCCGAGATGTGAAGTTGTCTGAGGAGAGATCCCTAT 453  
QY 450 caacgtgcccaacgaagacatgactctgttcaagtgtgcaaaagccagttccccccc 509  
DB 454 CACGACCTAATACCTACGGGCTCTGCTGCTCAGTGTCCAAAGACCAAGTTCCACCCCG 513  
QY 510 cccagctctgtaacatatacaagaattgtgacgacatgagcttcaagcattcc 569  
DB 514 CCAGATTGGAAAAACATTTTAAAGAGCTGTCTACAGACATCGATGG-TTTGTTCACTCT 572  
QY 570 ggaatgagatcctaagcagatgagcaaaaggggtgctgctgtcttaacgcgagtcg 629  
DB 573 GGCATAGGAGATTGTTCAGGGGTGGGCCGACAAAGGTCTCTCTCCCTCAAGCGCTCTCT 632  
QY 630 accgtgcgagccatcaagcgaactccacaaaggaagagcgtggagacctccacgac 689  
DB 633 ACTGTCCGCGCCCAACCAAGCCCAATTCCTCATTAAGAGAGGGGCTGGAGACAGTTCCAG 692  
QY 690 gctgtatcaagtgcgtgacgctcaacccgggaaggaagtccttcctcgtgtg-99gctc 748  
DB 693 GCGGTGCTGCTCGGTGATCAGAACCTGATGAGGCTCTCTCTCTCTCTCTGAGGCTC 752  
QY 749 ataagcccataagaag 764  
DB 753 TTATGCTCAGAGAGCG 768

## RESULT 15

BG866388

LOCUS 602185307F1 NCI\_CGAP\_SG2 Mus musculus CDNA clone IMAGE:4911553 5',  
DEFINITION mRNA sequence.

ACCESSION

BG866388

VERSION

BG866388.1

KEYWORDS

EST.

SOURCE

ORGANISM

614 bp mRNA linear EST 29-MAY-2001  
house mouse.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 614)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov

tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L14M10813 row: n column: 02  
High quality sequence stop: 614.

## FEATURES

source

1. 614  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4911553"  
/clone\_1db="NCI\_CGAP\_SG2"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: salivary gland; Vector: PCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.3 kb. Constructed by Life Technologies. Note: this is a NCI\_CGAP Library."  
BASE COUNT 137 a 169 c 168 g 140 t  
ORIGIN

Query Match 20.5%; Score 262.4; DB 10; Length 614;  
Best Local Similarity 71.5%; Pred. No. 5.8e-54;  
Matches 372; Conservative 0; Mismatches 146; Indels 2; Gaps 2;

QY 331 tttagctgtagagagagcgctcacacgcttcaacccagcgctgacgaattacagtt 390  
DB 97 TTGAGCTGAAGAAGAACACACAGGCTATCTCCGCCCGAGACAGTGTTCAT 136  
QY 391 cgaacagaatgtgtacatcacaagatgtgaagtagtgaattcaggccagacattac 450  
DB 157 GAGACCGATGTGCGACATCCGAGATGTGAAGTTGCTCATTTGGGAGAGATCCCTATC 216  
QY 451 acggtcccaacgaagacatgactgttcaagtgtgcaaaagccagttccccccc 510  
DB 217 ACCGACCTAATCAAGCTCACGGGCTGTCTCAGTGTCCAAAGACCAAGTTCCACCCCG 276  
QY 511 ccaagctctgtaacatatacaagaattgttaccgaacatgtagtgcctcaagcattc 570  
DB 277 CCAAGT-TGAAAAACATTTTAAAGAGCTGTCTACAGACATCGATGG-TTTGTTCACTCTG 334  
QY 571 gacatgagatcctaagcagatgagcaaaaggggtgctgctgtcctaagcagtgctga 630  
DB 335 GCCATGGGATTTGTGACGCTGGGCCCGACAAAGGTGCTCTCCCTCAAGCGGCTCTCA 394  
QY 631 ccgtgagggccatcaagcgaactcccaagaagagagcgtggagaccttcaacgagc 690  
DB 395 CTGTCCGCGCCCAACCAAGTTCCTAATGAAGAGAGGGGCTGGAGACAGTTCCAGGAG 454  
QY 691 ctgtgacgaagtgtgacgctcaacccgggaaggaagtccttcctcgttctgtggagc 750  
DB 455 CGGTGCTGCTCGGTGAATCAGAACCTGAGTGAGCTCTGCTCTCTCTCTGAGGCTCTT 514  
QY 751 acgccaataagagagagcaccatcagcaagaagaacgtccacatgtctcgaagctgttc 810  
DB 515 ATGCTCAGAAGAGGAGAGGTGATTCATTCAGAGGAAGCTCACCAATGTTCTCGACAGAC 574  
QY 811 atccatctccttctgtcgtcatcgtgtggttcctgtgtg 850  
DB 575 ACCCTTCGCGCTGCTGCTGATACAGAGGGTTCCTGGGATG 614



Search completed: August 22, 2002, 04:25:26  
Job time: 8136 sec

---





```

; APPLICANT: Ohki, Misao
; APPLICANT: Kikuchi, Kimiko
; APPLICANT: Miyoshi, Hiroyuki
; APPLICANT: Kozu, Tomoko
; TITLE OF INVENTION: No. 5580727e1 DNAs, Polypeptides Encoded Thereby
; TITLE OF INVENTION: and Methods for Detecting the DNAs and Polypeptides
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 810 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/244,189
; FILING DATE: 15-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 760-183P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 752 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-244-189-2

```

```

Query Match          5.6%; Score 89; DB 1; Length 752;
Best Local Similarity 21.8%; Pred. No. 0.67;
Matches 69; Conservative 36; Mismatches 82; Indels 130; Gaps 20;

QY 18 LPLGILLPQTLCSKLMITPKRLSSN---VEQKTSSPOLSVQLEEMANN-KKALDK 72
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 337 LP-LLORELLHCARLAKONPAQYLAHQEOLLIDASTSPVDSSELLDVNENGRKRPDR 395
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 73 IRAKATGAGFETWRRELAFFEKPYFKQLMSFVADERSRTVVPRAQVYSSSTEMCIIQ 132
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 396 TKEN---GF-----DREPLHSEHP-----SKRPCTIS 419
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 133 DVKVVILGDPYHGPNQAHGLCF---SVQKRPVPPP-----SLVNIYKELCTDIGF 181
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 420 P-----GQ--RYSFN--NGLSTQPNGLPHPTPPRPGHRIIDMAIAHNHYR-----DSY 463
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 182 KHPHGGLSG-----WAKGV---LLNAV-----L 204
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 464 RHPSHRDLRDNRPMGLHGTQOEIMIDRLDRMAEEMKMLDHLNCLIMDVEKTRSL 523
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 205 TV--RAHQANSHKDRGW-ETTTDAVYIKLSVNRGVVFLMGSTAHKKGATITDKRHNVL 261
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 524 TVLRRCEADERELNYWTRIRSDA-----BDL-----KKGCGSSSSHSRQ 564
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 262 QAVHPS--LSAHGGL 276
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 565 SPVNPDPVALDAHREFL 581
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 3
US-08-826-267-2
; Sequence 2, Application US/08826267
; Patent No. 5994070

```

```

; GENERAL INFORMATION:
; APPLICANT: Streuli, Michel
; TITLE OF INVENTION: No. 5994070e1 TRIO Molecules and Uses Related Thereo
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/826,267
; FILING DATE: 1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/014,214
; FILING DATE: 27 MARCH (1996)
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: DFN-010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2860 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-826-267-2

```

```

Query Match          5.1%; Score 81.5; DB 2; Length 2860;
Best Local Similarity 19.9%; Pred. No. 33;
Matches 59; Conservative 41; Mismatches 95; Indels 101; Gaps 13;

QY 12 ISSNRVLPGLLIPQTLCSKLMITPKRLSSNVE---QKTSSPOLSVQLEEMAKNK-- 66
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 684 ISSNK-----TPHNSSTNHETVLDQLEDAQSQWELFOEKRIKLE 724
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 67 -----KALDKIRAKATPAGFETWRRELAFFEKPYFKQLMSF-----VADERS 111
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 725 LFLHVRIFERDAIDILISL-----ESWNDELS-----QQWNPDTEDLTIAQRL 769
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 112 RHTVVPRAQVYSSSTEMCIIQDVKVVILGOD--PYNGPNOAHGLCFVQKRPVPPPSLVN 169
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 770 QH---HADKALITMNNL-----TFDVIHQGDLLQVNVQASGVLLCDRDVDMATRVQD 821
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 170 IY-----KELCTDIGEFKHPGH-----GLSGMAKOGVLLNA-VLTVRAN 209
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 822 LLEFLHEKQOELDLAEOHRRKHLLEQCYOLRHLDQAEVKQVLLGWIRNGESMLNAGLITASSL 881
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 210 QANSHKDRGWETTTDAVYIKLSVNRGVVFLMGSTYAHKKGATITDKRHNVLQANV 265
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 882 QEADQLOREHFOHAIEK-----THQSALQVOQKAEAMLDANH 920
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 4
US-08-631-607-6
; Sequence 6, Application US/08631607
; Patent No. 5767252
; GENERAL INFORMATION:
; APPLICANT: Morley et al, Paul
; TITLE OF INVENTION: NOVEL NEURONAL CELL GROWTH FACTOR
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:

```



APPLICANT: Guegler, Karl J.  
APPLICANT: Patterson, Chandra  
TITLE OF INVENTION: GROWTH-ASSOCIATED TRYPSIN-TYPE  
TITLE OF INVENTION: INHIBITOR HEAVY CHAIN PRECURSOR  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: Word Perfect 6.1/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/388,774  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/074,579  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Ceirone, Michael C  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PF-0505 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 946 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GENE BANK  
CLONE: g133985  
US-09-388-774-3

Query Match 5.0%; Score 80; DB 4; Length 946;  
Best Local Similarity 19.1%; Pred. No. 9.1;  
Matches 71; Conservative 42; Mismatches 113; Indels 146; Gaps 17;

QY 34 KITPKLR--SSNVEOKTSSPOLSVLEQER-----AKKKAALDKIRAKATPAGFGE 84  
DB 532 KFDPAKLDIESVITATSNQVLLETIAQMDLDQDFLSKDKHADPPETR-----K 582  
QY 85 TW-----RRELA--AEFEKPYKOLMSFVADERSRHTVPRPDQVSSSTEMDIO 132  
DB 583 LMAVLTNQLAERSLAPTAAKRRITRSIIQMSLD---HHITPILSTLIENAGDER 638  
QY 133 DVKAVILGDPYHGPNQAHGCFVSQKRPV-----PP 164  
DB 639 -----MLADAPROPDPSGSGALYYGSKVVPDSTPSMANPSTPIYISMLAOSQVLESTPP 693  
QY 165 PSLVNI-----KELCTDID-----GFKHPGHG 187  
DB 694 PHVMRVENDPFIILYLPKSGKNICFNIDSEPGILNLVSDPSGIVVNGQLVGAKKPPNG 753  
QY 188 DLSGW-AKQGVLLINAVLTAHQAHSKDRGWTFTDAVTKW-----LSVNR 234  
DB 754 KLSYTFEKLGFYQSEDIKIEISTETITLSHGSSTFS--LSWSDTAQVTVNQRVOISYK 810  
QY 235 EGVVFLMGSAHKKGATIDKRH-----HYLAHVPSPLSAHNGFLGC-----KHPSK 283  
DB 811 EKVYTI-----TLDEKMSFVLLHRVWKHP-----VNVDFLGITYIPPNKSPK 855  
QY 284 ANGLIKLSTEP 295

DB 856 AHGLIGFMDEP 867  
RESULT 7  
US-08-631-607-7  
Sequence 7, Application US/08631607  
Patent No. 5767252  
GENERAL INFORMATION:  
APPLICANT: Morley et al, Paul  
TITLE OF INVENTION: NOVEL NEURONAL CELL GROWTH FACTOR  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/631,607  
FILING DATE: 08-Apr-1996  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07265/086001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 430 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-631-607-7

Query Match 4.9%; Score 79; DB 1; Length 430;  
Best Local Similarity 23.9%; Pred. No. 3.6;  
Matches 57; Conservative 25; Mismatches 94; Indels 62; Gaps 12;

QY 56 VEQLERNAKKKAALD---KIRAKATPAGGFTWRREL---AAEFKPYFKOLMSFYA- 107  
DB 123 VEQLSRSLQTLKRLSEHQLRANVSAGCPDGFREVLDQRLGELE---RQLLRKYAE 178  
QY 108 --DERS-RHYVYPPADQVYSTEMCDIQDVKVVILGDPYHGPQAHGICFSYQKRPV- 163  
DB 179 LEDERSLHNETSARHKTSTLNAALQRYTELRSNAFESPD-----AKVSLPLKTN 233  
QY 164 -----PPSLVNIYK-ELCTDIDGFKHPGHGDLGMAKOG-----VLLLN 201  
DB 234 YLYGKIKKTLPELAYFTICLMLRSSASPGIGTPSYAVPGANELLILEMGNPIELLIN 293  
QY 202 ---AVLTVRAHQANSH-----KDRGWETFTDA-----VIKLSVNRGCVFL 240  
DB 294 DKVAQDLPLFVSDGKWMHHCVTWTRDGMWMAFQDEKLGTCGENLAPMHPINPGVLLI 351

RESULT 8  
US-08-568-459A-12  
Sequence 12, Application US/08568459A  
Patent No. 5849306  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan

APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobb Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/568,459A  
FILING DATE: 07-DEC-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelson, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2710 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium falciparum  
US-08-568-459A-12

Query Match 4.9%; Score 78; DB 2; Length 2710;

Best Local Similarity 22.8%; Pred. No. 74;

Matches 71; Conservative 34; Mismatches 129; Indels 78; Gaps 15;

29 FSKLMKTFPKKLSSNVQKTSPOLSVEQLERMAK-----NKKKALDKIRAKATPAG 81  
1496 FSKDGSKSPSGL-SROEMWKTNGPEIWKGMICALTKYVYDTPDKRKRIKNDYSIDKVNQSQ 1554  
82 FGETWRRELAEEFEKPYFKOLM-----SFVADERSRHTVYPPADQVYSTEMCDIODVKY 136  
1555 NGNPSLEEFAN--KPPFLMWMIEMGEPECAERQKKNIIKDCNENLSNQCN----- 1605  
137 VILGDPYHGPNOAHGLCFVYOKVPPPPSLVNIYKELCTIDIGF-----KHGEGDLSG 191  
1606 -----DAKHRCNOA--CRAVYOEVY-----NKKKEFSQTNNFVLKANYQPDDEPKG 1651  
192 WA-KOGVLLANVLTVRAHQANSHK-----DRGWETFLDAVIKMLSVNREGVFLMGSYA 246  
1652 YEYKDG-----VPIQGNETLLOKCDNNKCSMDGNV--LSVSPKEKPF---GKTA 1697  
247 HKKGATIDRRKHHVLAQVHSPSLSAHRG-----FLGCKHFSKANGILK 289  
1698 HKYPERKCDYQGNKHPVSIPIPPPPVQPPQPEAPVTVVDVCSIVKTLFQDTNNFSDAGC-LK 1756  
290 LSGTEPIINRAL 301  
1757 YGKTAPSSMKCI 1768

RESULT 9

US-08-487-826B-12  
Sequence 12, Application US/08487826B  
Patent No. 5993827  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chiltons, Chetan  
APPLICANT: Peterson, David S.  
APPLICANT: Miller, Louis H.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobb Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelson, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2710 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium falciparum  
US-08-487-826B-12

Query Match 4.9%; Score 78; DB 2; Length 2710;

Best Local Similarity 22.8%; Pred. No. 74;

Matches 71; Conservative 34; Mismatches 129; Indels 78; Gaps 15;

29 FSKLMKTFPKKLSSNVQKTSPOLSVEQLERMAK-----NKKKALDKIRAKATPAG 81  
1496 FSKDGSKSPSGL-SROEMWKTNGPEIWKGMICALTKYVYDTPDKRKRIKNDYSIDKVNQSQ 1554  
82 FGETWRRELAEEFEKPYFKOLM-----SFVADERSRHTVYPPADQVYSTEMCDIODVKY 136  
1555 NGNPSLEEFAN--KPPFLMWMIEMGEPECAERQKKNIIKDCNENLSNQCN----- 1605  
137 VILGDPYHGPNOAHGLCFVYOKVPPPPSLVNIYKELCTIDIGF-----KHGEGDLSG 191  
1606 -----DAKHRCNOA--CRAVYOEVY-----NKKKEFSQTNNFVLKANYQPDDEPKG 1651  
192 WA-KOGVLLANVLTVRAHQANSHK-----DRGWETFLDAVIKMLSVNREGVFLMGSYA 246  
1652 YEYKDG-----VPIQGNETLLOKCDNNKCSMDGNV--LSVSPKEKPF---GKTA 1697  
247 HKKGATIDRRKHHVLAQVHSPSLSAHRG-----FLGCKHFSKANGILK 289  
1698 HKYPERKCDYQGNKHPVSIPIPPPPVQPPQPEAPVTVVDVCSIVKTLFQDTNNFSDAGC-LK 1756





```

US-08-066-167-1
: Sequence 1, Application US/08066167
: Patent No. 5618541
: GENERAL INFORMATION:
: APPLICANT: QUENTIN-MILLET, Marie-Jose
: TITLE OF INVENTION: VACCINE AGAINST NEISSERIA MENINGITIDIS
: TITLE OF INVENTION: INFECTIONS
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Burns, Doane, Swecker & Mathis
: STREET: George Mason Bldg., Washington & Prince Sts.
: CITY: Alexandria
: STATE: Virginia
: COUNTRY: United States
: ZIP: 22313-1404
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/066,167
: FILING DATE: 02-JUN-1993
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 91 12177
: FILING DATE: 03-OCT-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Crane-Feury, Sharon E
: REGISTRATION NUMBER: 36,113
: REFERENCE/DOCKET NUMBER: 016100-002
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 836-6620
: TELEFAX: (703) 836-2021
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 579 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: ORIGINAL SOURCE:
: ORGANISM: Neisseria meningitidis 2394 subunit Tbp2.
US-08-066-167-1

Query Match 4.8%, Score 77.5; DB 1; Length 579;
Best Local Similarity 26.6%, Pred. No. 8.1;
Matches 45; Conservative 26; Mismatches 69; Indels 29; Gaps 9

QY 76 KATAGPFTMRRLAAFEKPYFQOLMSFVADERSR-----HTYPPADQYYSTEMCD 130
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 80 RGEPSFSE--RDEL--EKRRSSSELISKWEDGGSRYVGYTNTFYRSGYYLKNKNMID 135
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 131 IQDVKVVLTGDDPYHGPNQAGLCFSYOKRPVPPPSLVNIYK---ELCD-IDGFKHPGH 186
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 136 IKN-NIVLFGPDGY-----LTYKKGKPEKSLPSEKITYGTMDYVDANEMKQRFEE 186
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 187 GDLSGMAQGV--LLNAVLTIVRAHQANSKDRGET-----FTDAVTK 228
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 187 GSAAGGDKSGALSALEEGVLRNQAEASGHRDFGMTSEFEVDVSDKTIK 235
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
: US-08-449-733-1
: Sequence 1, Application US/08449733
: Patent No. 5928650
: GENERAL INFORMATION:
: APPLICANT: QUENTIN-MILLET, Marie-Jose
: APPLICANT: LISSOLO, Ling
: TITLE OF INVENTION: SUBUNIT VACCINE FOR NEISSERIA
: TITLE OF INVENTION: MENINGITIDIS INFECTIONS AND CORRESPONDING PURIFIED
: TITLE OF INVENTION: SUBUNTS

```

```

NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESSES:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,733
FILING DATE: 25-MAY-1995
CLASSIFICATION: 424

PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 12176
FILING DATE: 03-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/064,174
FILING DATE: 25-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm K.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 016100-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703/836-6620
TELEFAX: 703/836-2021
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 579 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-449-733-1

Query Match 4.8%; Score 77.5; DB 2; Length 579;
Best Local Similarity 26.6%; Pred. No. 8.1;
Matches 45; Conservative 26; Mismatches 69; Indels 29; Gaps 9

QY 76 KATPAGFETTRRLAEFEKPYEQLMSEVADERSR-----HTYVPADQVYSSTEMCD 130
D 80 KGEPSFSE--RDEL--EKRRGSSSLIESKMGDSRVVGYTNFTYVRSGYVLLKNMND 135
QY 131 IODYKVVVLIGDDPYGPNQAHGLCFSVQKVPYRPPSLVNIYK---ELCPD-IDGFKHPGH 186
D 136 IKN-NIVLEFGPDGY-----LYYKKEPFSKELPSEKITYTKGTWDYITDAMEKORFESL 186
QY 187 GDLGSMAGQVLL--LLNAVLTVYRAHQANSHKDRGME-----FTDAVIK 228
D 187 GSAAGGDKSGALSLALEEGVLRMQAEASGHTDFGMTSEFEVDFSDKTIK 235

RESULT 14
US-08-487-890A-97
Sequence 97, Application US/08487890A
Patent No. 5708149
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena
APPLICANT: Harkness, Robin
APPLICANT: Schryvers, Anthony
APPLICANT: Chong, Pele
APPLICANT: Gray-Owen, Scott
APPLICANT: Yang, Yan-Ping
APPLICANT: Murlin, Andrew
APPLICANT: Klein, Michel
TITLE OF INVENTION: Transferrin Receptor Genes
NUMBER OF SEQUENCES: 147
CORRESPONDENCE ADDRESSES:

```

ADDRESSEE: Sim & McBurney  
STREET: 6th Floor, 330 University Avenue  
City: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,890A  
FILING DATE: 07-JUN-1993  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/175,116  
FILING DATE: 29-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/148,968  
FILING DATE: 08-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-466 MIS:jib  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 97:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 598 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-487-890A-97

Query Match 4.8%; Score 77.5; DB 1; Length 598;  
Best Local Similarity 26.6%; Pred. No. 8.5;  
Matches 45; Conservative 26; Mismatches 69; Indels 29; Gaps 9;

QY 76 KATPAGGEWRRRLAEEFKPYRKQMSFVADRSR-----HTVPPADQVYSTKCD 130  
DB 100 RGEPNSESE--RDEL--EKRRGSELIESKWEQDSRVGVTNFTYVRSQVYVYLNKNID 155  
QY 131 IODKVVITLGDPPYHGNQAHGLCFVSQKPYPPPSLVNITYK---ELCTD-IDGFKHPGH 186  
DB 156 IKN-NIVLFEPDGY-----LYYKGEPSKELPSEKITKGTWDVYTDAMEKQREGL 206  
QY 187 GDLGMAKQGV--LLNAVLTVRAHQANSKDRGMET----FTDAVIK 228  
DB 207 GSAAGDGSKGLSLAEGLVLRNQAEASSGHTDFGMTSEFVDFSDKTIK 255

RESULT 15  
US-08-478-435-97  
Sequence 97, Application US/08478435  
Patent No. 5923323  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena  
APPLICANT: Hartness, Robin  
APPLICANT: Schuyvers, Anthony  
APPLICANT: Chong, Pele  
APPLICANT: Gray-Owen, Scott  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Murdin, Andrew  
APPLICANT: Klein, Michel  
TITLE OF INVENTION: Transferrin Receptor Genes  
NUMBER OF SEQUENCES: 147  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Slim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto

STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/478,435  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/337,483  
FILING DATE: 08-NOV-1994  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/175,116  
FILING DATE: 29-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/148,968  
FILING DATE: 08-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-462 MIS:vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 97:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 598 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-478-435-97

Query Match 4.8%; Score 77.5; DB 2; Length 598;  
Best Local Similarity 26.6%; Pred. No. 8.5;  
Matches 45; Conservative 26; Mismatches 69; Indels 29; Gaps 9;

QY 76 KATPAGGEWRRRLAEEFKPYRKQMSFVADRSR-----HTVPPADQVYSTKCD 130  
DB 100 RGEPNSESE--RDEL--EKRRGSELIESKWEQDSRVGVTNFTYVRSQVYVYLNKNID 155  
QY 131 IODKVVITLGDPPYHGNQAHGLCFVSQKPYPPPSLVNITYK---ELCTD-IDGFKHPGH 186  
DB 156 IKN-NIVLFEPDGY-----LYYKGEPSKELPSEKITKGTWDVYTDAMEKQREGL 206  
QY 187 GDLGMAKQGV--LLNAVLTVRAHQANSKDRGMET----FTDAVIK 228  
DB 207 GSAAGDGSKGLSLAEGLVLRNQAEASSGHTDFGMTSEFVDFSDKTIK 255

Search completed: August 21, 2002, 10:24:56  
Job time: 258 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: August 21, 2002, 10:22:13 ; Search time 17.5 Seconds

(without alignments)  
1652.737 Million cell updates/sec

Title: US-09-758-017a-2

Sequence: 1 MLEFKGLGCRICISSNRVLPG.....SRANGLKLKSGTEPIWMRAL 301

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summariesDatabase : PIR.71:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1032	64.4	304	1 A60472	uracil-DNA glycosylase
2	705.5	44.0	228	2 AH0329	probable uracil-DN
3	702.5	43.8	226	2 C82086	uracil-DNA glycosylase
4	683.5	42.6	229	2 DB5904	uracil-DNA glycosylase
5	681.5	42.5	229	1 DGEUC	uracil-DNA glycosylase
6	679.5	42.4	229	2 AH0830	uracil-DNA glycosylase
7	659.5	41.1	218	2 F91059	uracil-DNA glycosylase
8	648.5	40.5	219	2 H81107	uracil-DNA glycosylase
9	635.5	39.6	247	2 A64043	uracil-DNA glycosylase
10	635.5	39.6	247	2 F75486	uracil-DNA N-glyco
11	634.5	39.6	219	2 B81908	probable uracil-DN
12	598	37.3	237	2 AE2878	uracil-DNA glycosylase
13	598	37.3	241	2 G97654	uracil-DNA glycosylase
14	591	36.9	256	2 H82525	uracil-DNA glycosylase
15	586.5	36.6	231	2 C83551	uracil-DNA glycosylase
16	584	36.4	217	2 I36145	uracil-DNA glycosylase
17	572.5	35.7	224	2 B84131	uracil-DNA glycosylase
18	570.5	35.6	225	2 S39712	uracil-DNA glycosylase
19	569	35.5	231	2 F81424	uracil-DNA glycosylase
20	561	35.0	222	2 T40846	uracil-DNA glycosylase
21	559.5	34.9	224	2 AC1228	uracil-DNA glycosylase
22	559.5	34.9	224	2 AE1581	uracil-DNA glycosylase
23	557	34.7	218	2 G89826	uracil-DNA glycosylase
24	552	34.4	229	2 B81652	uracil-DNA glycosylase
25	551	34.4	220	2 DB4951	uracil-DNA glycosylase
26	548	34.2	229	2 C71493	probable uracil DN
27	546.5	34.1	221	2 AG1123	uracil-DNA glycosylase
28	540	33.7	359	1 A31425	uracil-DNA glycosylase
29	539.5	33.7	236	2 C86587	uracil DNA glycosylase

30	539.5	33.7	236	2 H72035	uracil-DNA glycosylase
31	536.5	33.5	221	2 A11483	uracil-DNA glycosylase
32	535.5	33.4	217	2 E95135	uracil-DNA glycosylase
33	533.5	33.3	217	2 S13591	uracil-DNA glycosylase
34	533.5	33.3	222	2 G98003	DNA-uracil glycosylase
35	496.5	31.0	223	2 E70106	uracil-DNA glycosylase
36	474	29.6	314	2 T42604	probable uracil-DN
37	472.5	29.5	312	1 DGBE83	uracil-DNA glycosylase
38	472	29.4	252	2 T03140	uracil-DNA glycosylase
39	463.5	28.9	281	2 S33023	probable uracil-DN
40	457	28.5	255	2 S55641	uracil-DNA glycosylase
41	454.5	28.4	219	2 B86654	uracil-DNA glycosylase
42	451	28.1	252	2 T42961	probable uracil-DN
43	446.5	27.9	212	2 G62877	uracil-DNA glycosylase
44	445	27.8	252	1 DGBE88	uracil-DNA glycosylase
45	429.5	26.8	250	1 DGBE85	uracil-DNA glycosylase

## ALIGNMENTS

```
RESULT 1
A60472
uracil-DNA glycosylase (PC 3.2.2.-) precursor [validated] - human
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 08-Dec-2000
C:Accession: S05964; A60472
R:Olson, L.C.; Aasland, R.; Wiltner, C.U.; Krokan, H.E.; Helland, D.E.
EMBO J. 8, 3121-3125, 1989
A:Title: Molecular cloning of human uracil-DNA glycosylase, a highly conserved DNA re
A:Reference number: S05964; MUID:90059899
A:Accession: S05964
A:Molecule type: mRNA
A:Residues: 1-304 <OLS>
A:Cross-references: EMBL:X15653; NID:g37598; PID:CAA33679.1; PID:g37599
R:Wiltner, C.U.; Bauw, G.; Krokan, H.E.
Biochemistry 28, 780-784, 1989
A:Title: Purification and determination of the NH-2-terminal amino acid sequence of u
A:Reference number: A60472; MUID:89229080
A:Accession: A60472
A:Molecule type: Protein
A:Residues: 78-93, 'X', 95-104 <WIT>
C:Genetics:
A:Gene: GDB:UNG, DGU, UDG
A:Cross-references: GDB:119844; OMIM:191525
A:Map position: 12pter-12qter
C:Superfamily: uracil-DNA glycosylase
C:Keywords: DNA repair; glycosidase; hydrolase; mitochondrion
F:1-77/Domains: transic peptide (mitochondrion) #status predicted <TNP>
F:78-304/Product: uracil-DNA glycosylase #status experimental <MAT>

Query Match 64.4%; Score 1032; DB 1; Length 304;
Best Local Similarity 67.4%; Pred. No. 6-3e-81;
Matches 194; Conservative 32; Mismatches 56; Indels 6; Gaps 3;

OY 19 PG---LLPQLCFSKLMKTPKLRSSNVEQKT--SSFQLSVEQDLERNAKKKKAAALDKI 73
|| | : || | | | | | | | | | | | | | | | | | | | | | | | | |
DB 18 PKGKPLQLLSRLCGDHLQAIIPAKKAPAGQEPGTPPSSP-LSAQEDRIQRKKAAALRL 76
74 RAKATPAGGEGTWRKRELAEEFEKPYFKOLMSFYADERSNHYTPPADQVYSSSTEMCDIOD 133
77 AARNPVGVEGSMKKRLDSEFGKPYFIKMGFVAERKHYTVYPPHOVFTWTQCDIKD 136
OY 134 VVVVILGDPYHGPQNAHGLCFSVOKPVPVPSLVNIYKELCTDIDGFKHPHGLDSGMA 193
|||||GDPYHGPQNAHGLCFSVOKPVPVPSLVNIYKELCTDIDGFKHPHGLDSGMA 193
DB 137 VVVVILGDPYHGPQNAHGLCFSVOKPVPVPSLVNIYKELCTDIDGFKHPHGLDSGMA 196
OY 194 KGGVLLNNAVLTVRHQANSKDRGMEFTDAVIKMLSVNREGVYVFLMGSAHKKGATV 253
|||||GDPYHGPQNAHGLCFSVOKPVPVPSLVNIYKELCTDIDGFKHPHGLDSGMA 253
DB 197 KGGVLLNNAVLTVRHQANSKDRGMEFTDAVIKMLSVNREGVYVFLMGSAHKKGATV 256
OY 254 DKRRHNVLDQAVHPSPLSAHRGFLGCKHPSKANGLKLKSTGEPINMRAL 301
```

Db 257 DRKRHHVLOTAAHPSPLSVYRGFGCRHPSKTNELLQKSGKKPIDMKEL 304

RESULT 2

AH0329

Probable uracil-DNA glycosylase (EC 3.2.2.-) [Imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis

C>Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 27-Nov-2001

C:Accession: AH0329

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tiltball, R.W.; Holden, M.T.G.; Prentice, M.B.

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,

Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AH0329

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-228 <KUR>

C:Cross-references: GB:AL590842; PIDN:CAC92943.1; PID:g15980682; GSPDB:GN00175

C:Genetics:

A:Gene: ung

C:Superfamily: uracil-DNA glycosylase

C:Keywords: glycosidase; hydrolase

Query Match 44.0%; Score 705.5; DB 2; Length 228;  
Best Local Similarity 60.2%; Pred. No. 4.6e-53;  
Matches 130; Conservative 34; Mismatches 51; Indels 1; Gaps 1;

Db 85 TWRELAEFKPYFKOLMSFVADERSRHTVVPADQVYSTKCDIDYKVVILGDP 143

Y 144 YHGRNOAHGLCFYQKRPVPPPSLVNLYKELCTDIDGFKHGHGDLGSMARQGVLLNAV 203

Db 66 YHGRNOAHGLCFYQKRPVPPPSLVNLYKELCTDIDGFKHGHGDLGSMARQGVLLNAV 125

Y 204 LTVRAHQANSKRDGMEFTDAVTKLSVNRGVFLMGSAVANKKATIDRKHHVLOA 263

Db 126 LTVRAHQANSKRDGMEFTDAVTKLSVNRGVFLMGSAVANKKATIDRKHHVLOA 185

Y 264 VHPSPLSAHRGFLGCKHPSKANGLIKLSGTEPIMR 299

Db 186 PHPSPLSAHRGFLGCKHPSKANGLIKLSGTEPIMR 221

RESULT 3

uracil-DNA glycosylase VC2359 [Imported] - Vibrio cholerae (strain N16961 serogroup O1)

C:Species: Vibrio cholerae

C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C:Accession: C82086

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;

Chadson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, H.

1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833

A:Accession: C82086

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-226 <HEIR>

C:Cross-references: GB:AE004306; GB:AE003852; NID:g9656924; PIDN:AAF95502.1; GSPDB:GN001

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC2359

A:Map position: 1

C:Superfamily: uracil-DNA glycosylase

Query Match 43.8%; Score 702.5; DB 2; Length 226;

Best Local Similarity 61.1%; Pred. No. 8.2e-53;  
Matches 132; Conservative 25; Mismatches 58; Indels 1; Gaps 1;

Db 85 TWRELAEFKPYFKOLMSFVADERSRHTVVPADQVYSTKCDIDYKVVILGDP 143

Y 144 YHGRNOAHGLCFYQKRPVPPPSLVNLYKELCTDIDGFKHGHGDLGSMARQGVLLNAV 203

Db 66 YHGRNOAHGLCFYQKRPVPPPSLVNLYKELCTDIDGFKHGHGDLGSMARQGVLLNAV 125

Y 204 LTVRAHQANSKRDGMEFTDAVTKLSVNRGVFLMGSAVANKKATIDRKHHVLOA 263

Db 126 LTVRAHQANSKRDGMEFTDAVTKLSVNRGVFLMGSAVANKKATIDRKHHVLOA 185

Y 264 VHPSPLSAHRGFLGCKHPSKANGLIKLSGTEPIMR 299

Db 186 PHPSPLSAHRGFLGCKHPSKANGLIKLSGTEPIMR 221

RESULT 4

uracil-DNA-glycosylase [Imported] - Escherichia coli (strain O157:H7, substrain EDL93

C:Species: Escherichia coli

C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C:Accession: D85904

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May

11ler, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: D85904

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-229 <STO>

C:Cross-references: GB:AE005174; NID:g12517001; PIDN:AAG57696.1; GSPDB:GN00145; UMGF:

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: ung

C:Superfamily: uracil-DNA glycosylase

Query Match 42.6%; Score 683.5; DB 2; Length 229;

Best Local Similarity 60.9%; Pred. No. 3.6e-51;

Matches 131; Conservative 24; Mismatches 59; Indels 1; Gaps 1;

Db 85 TWRELAEFKPYFKOLMSFVADERSRHTVVPADQVYSTKCDIDYKVVILGDP 143

Y 144 YHGRNOAHGLCFYQKRPVPPPSLVNLYKELCTDIDGFKHGHGDLGSMARQGVLLNAV 203

Db 66 YHGRNOAHGLCFYQKRPVPPPSLVNLYKELCTDIDGFKHGHGDLGSMARQGVLLNAV 125

Y 204 LTVRAHQANSKRDGMEFTDAVTKLSVNRGVFLMGSAVANKKATIDRKHHVLOA 263

Db 126 LTVRAHQANSKRDGMEFTDAVTKLSVNRGVFLMGSAVANKKATIDRKHHVLOA 185

Y 264 VHPSPLSAHRGFLGCKHPSKANGLIKLSGTEPIMR 298

Db 186 PHPSPLSAHRGFLGCKHPSKANGLIKLSGTEPIMR 220

RESULT 5

uracil-DNA glycosylase (EC 3.2.2.-) - Escherichia coli

C:Species: Escherichia coli

C>Date: 31-Dec-1989 #sequence\_revision 31-Dec-1989 #text\_change 16-Jun-2000

C:Accession: A28175; C65036

R:Varshney, U.; Hutcheon, T.; van de Sande, J.H.

J. Biol. Chem. 263, 7776-7784, 1988

A:Title: Sequence analysis, expression, and conservation of Escherichia coli uracil D

A:Reference number: A28175; MUID:88227981

A:Accession: A28175  
A:Molecule type: DNA  
A:Residues: 1-229 <VAR>  
A:Cross-references: GB:D64044; NID:q987635; PIND:BA10923.1; PID:q987651  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CC  
.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A>Title: The complete genome sequence of *Escherichia coli* K-12.  
A:Reference number: A64720; MUID:97426617  
A:Accession: C65036  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-229 <BLAT>  
A:Cross-references: GB:A000344; GB:U00096; NID:q1788927; PIND:AACT5633.1; PID:q1788934  
A:Experimental source: strain K-12, substrain M61655  
A:Comment: This enzyme catalyzes the hydrolysis of the deoxyribose N-glycosidic bond to  
on.  
C:Genetics:  
A:Gene: ung  
A:Map position: 56 min  
C:Superfamily: uracil-DNA glycosylase  
C:Keywords: DNA repair; glycosidase; hydrolase

```

Query Match          42.5%: Score 681.5; DB: 1; Length 229;
Best Local Similarity 60.9%: Pred. No. 5.3e-51;
Matches 131; Conservative 24; Mismatches 59; Indels 1; Gaps 1;

Oy 85 TWRRLEAEFEKPYFKQMLSFVADER-SRHVYVPADQVYSTECMDIQDVKVYLLGDP 143
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 6 TWHVDYLAEEKQPFYFLNTLQTVASERQSGVITYPQKQDVFNARFPELTGDVKVYLLGDP 65

Oy 144 YHGPQWAGLGEFSYOKPYRPPPSLVNRYKELCTIDGCKHKGHGDLSMAQGVYLLNAV 203
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 66 YHGPQWAGLGEFSVAPRGAIPLPSLMTKLELNTIPGTRPHNGILSMARQGVYLLQTV 125

Oy 204 LTVRAHQANSKDKRGWETETDAVIKWLTSVNEGYYVLLMGSYAHKKGATIDRKRHHVLA 263
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 126 LTVRAQGHSHASLGMETFTDKVISLINOHREGVYVLLMGSHAKKGA1IDKQRIHVLA 185

Oy 264 VHPSPLSAHRGFLGCKHPSKANGLLKLSGTEPIINW 298
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 186 PHPSPLSAHRGFPCCNHFVLANQWLEORGEPIDW 220

RESULT 6
AH0830
uracil-DNA glycosylase (EC 3.2.2.-) [imported] - Salmonella enterica subsp. enterica serovar typhimurium
C:Species: Salmonella enterica subsp. enterica serovar typhimurium
A:Note: this species has also been called Salmonella typhimurium
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C:Accession: AH0830
R:Packhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
  T.; Connerston, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
  S.; Moule, S.; O'Gaora, P.
  Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
  A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar
  A:Reference number: AB0302; PMID:11677608
A:Accession: AH0830
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-229 <PAR>
A:Cross-references: GB:AL51382; PIDD:CAD05831.1; PIDD:G16503806; GSPDB:GN00176
C:Genetics:
A:Gene: STY2840
C:Superfamily: uracil-DNA glycosylase
C:Keywords: glycosidase; hydrolase

```

Query Match	42.48;	Score 679.5;	DB 2;	Length 229;
Best Local Similarity	60.98;	Pred. No. 7.9e-51;		
Matches 131; Conservative	24;	Mismatches 59;		
				Indels 1; Gaps 1;

OY	85	TWRRLAEFEKPYRKOJMSFVADER--SRHNVUPRADVOYSTCEMDIOAKVVILLGDP	143
		: : : : :   : :         : : : : :   : : : : :	
Dd	6	TMDVLAEKQOOPYINTLHTVAGEROSGITVYPPOKVFNARFTELGSDVKVVIILGDP	65
OY	144	YHGPNQAHGLCFSVOKVPVPPPSLVNIYKELCTDIDGFKHBGHDDLGSMAKOVYLLNAV	203
		: : : : :   : : : : :   : : : : :   : : : : :	
Dd	66	YHGPGQAHGLAFSPVRGIAPPPLSLVMYKLELASIPGVPRAHGYLESMAHQVLLNTV	125
OY	204	LTVRAHQANSKDKRCHEFTTDAVTKMLSVNREGVYFLMSGVIAHKKATDIRKHNYLOA	263
		: :   :         : : : :         : :   : :   : :	
Dd	126	LTVRGAQHSHASLSQWEFFTDKVISLINQHRGGVFLLMSGSHAQKKAIIIDPORNHLIKA	185
OY	264	VHPSPLSAHRGFLGCCKHFPSKANGILKLXSGTBEPIN 298	
		: :   : :   : :   : :	
Dd	186	PHPSPLSAHRGFGCGNHFAALTQWLEOHGEKTIIDW 220	

```

RESULT 7
F91059
uracil-DNA-glycosylase ECS3446 [imported] - Escherichia coli (strain O157:H7, substra
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence-revision 18-Jul-2001 #text-change 03-Aug-2001
C:Accession: F91059
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
gasawara, N.; Yasunaga, T.; Kuhnara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
Gaster Res. 8, 11-22, 2001
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g
A:Reference number: A99629; MUID:21156231; PMID:1158796
A:Accession: F91059
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 17218 <HAY>
A:Cross-references: GB:BA000007; PIDN:BAR36869.1; PID:q13362917; GSPDB:GN00154
A:Experimental source: Strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECS3446
C:Superfamily: uracil-DNA glycosylase

```

	Query Match	41.1%;	Score 659.5;	DB 2;	Length 218;	
	Blast Local Similarity	61.7%;	Pred. No.3.8e-49;			
	Matches 127; Conservative	23;	Mismatches 53;	Indels 1;	Gaps 1;	
OY	TWRRELAERFEKPYFKQMLSFVADER-SRNHYVPRAPOVYSSTENCIDIODYKVYLIGDPR	143				
Dd	6 TWHDVLAEKKOORPFLNTLQTIVASERGSGVIITYPKQDVFNAPFRTLEGGDKVVILLGDR	65				
OY	144 YHGRPAAGGLCFESVOKRPPRRPSLVNITYKELCSTDIDGKNRGHODLSMAAQGYLLNAV	203				
Dd	66 YHGPGGAAGLGAFSVNPGIATPPSLNMKYELENTIPGETFRNHGSLEGSAWAGVLLNTPV	125				
OY	204 LTVRHAQNSKHNDRWMEFTDAVILKWLSVNREGGVVELLMGSIYAHHKGATDIRKHNHYLOA	263				
Dd	126 LTVRGAQHSHASHLSGWETFDTXVISLINQRHEGVVELMGSHACKKKAALIDKOHNHYLKA	185				
OY	264 VHPSPLSAHRGFGELCKHFPSKANGLIK	289				
Dd	186 PHPSPLSAHRGFFCGCNHFFVLNQWLDE	211				

RESULT 8  
H81107  
uracil-DNA glycosylase NMB1222 [imported] - *Neisseria meningitidis* (strain MC58 serog C);Species: *Neisseria meningitidis*  
C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C;Accession: H81107  
R;Retrieved: H.;Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, R.;Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Pougherty, B. ri, H.; Qin, H.; Vamthyan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M. Science 287, 1809-1815, 2000  
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappelli, R.;Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.

**RESULT**

<b>Db</b> 182 APHPSPLSAHHGFCCGRHFSKTNSTYLESHGICPDMQ F75486 uracil-DNA N-glycosylase - Deinococcus radiodurans (strain R1) C:Species: Deinococcus radiodurans C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000 C:Accession: F75486 R.White, O.; Eissen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J. M., Shen, M.; Vamathevan, J.T.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.C. S.: Smith, H.O.; Venter, J.C.; Fraser, C.M. Science 286, 1571-1577, 1999 A>Title: Genome sequence of the radioreistant bacterium Deinococcus radiodurans R1. A:Reference number: AF5250; MUID:20036896 A:Accession: F75486 A:Molecule type: DNA A:Status: preliminary A:Residues: 1-247 <MNT> A:CROSS-references: GB:AEO01926; GB:AEO00513; NID:g6458395; PIDN:AAFL10269.1; PID:g6458395 A:Experimental source: strain R1 C:Genetics: A:Gene: DR0689 A:Map position: 1 C:Superfamily: uracil-DNA glycosylase	<b>Query Match</b> Best Local Similarity      39.6%; Score 635.5; DB 2; Length 247; 56.1%; Pred. No. 5.2e-47; Matches 124; Conservative 27; Mismatches 69; Indels 1; Gaps 1;
--	---

```

OY   79  PAGCEGTTRRELAEFEKPYPYKOLMSFYADESRRTITVPADQVYSSTEMCDIODVKVI 138
          || | | | | | | | : : : : : : : : : : : : : : : : : : : :
DB   20  PANLPEDMQEALLPEFSAPYFHEILDFLROEKETIIPPAADVFNALRYTPPLGEVKLI  79
          || | | | | | | | | | | | | | | | | | | | | | | | | | |
OY   139 LGODPYHRPNAHGICEFVRKVPRPPPSLVNIYTKELCIDIGFKNPBGDLSGNAGOSVL 198
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB    80  LGDDYHBPNOAHGSSEFVSRRGVPRPRLNRIKELTDIEDGFVAARKGYLRSMAGSGL 139
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
OY   199 LLNAVLTTRAQAONSHKDRGWETTDAVIKMLSVNRREGVILLMGSYAHKKGATIDRKRH 258
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB   140 LLNAVLTTRAQAONSHOKGMENHFIDAIVKANAKKEENVYLIMGSYARKKKLITGKNH 199
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
OY   259 HVLAQVHPSPLSAHRGLGCCKHSKANGLLKLGSTEPIINMR 299
          ||::||| | | | | | | | | | | | | | | | | | | | | |
DB    200 VIIESGHSPSL-S-EQYFFGTRPFPSKTNEALKRGAGPVEMQ 239
          ||::||| | | | | | | | | | | | | | | | | | | | | |

```

**RESULT**

<b>B81908</b> probable uracil-DNA glycosylase (EC 3.2.2.-) NMA1384 [imported] - Neisseria meningitidis C:Species: Neisseria meningitidis C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 24-May-2001 C:Accession: B81908 Parkinson, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mc Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre Nature 404, 502-506, 2000 A>Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z249J A:Reference number: AB1775; MUID:20222556 A:Accession: B81908 A:Status: preliminary A:Molecule type: DNA A:Residues: 1-219 <PAR> A:CROSS-references: GB:AL162755; GB:AL157959; NID:g7379742; PIDN:CAB84630.1; PID:g738 A:Experimental source: serogroup A, strain Z2491 C:Genetics: A:Gene: ung; NMA1384 C:Superfamily: uracil-DNA glycosylase C:Keywords: glycosidase; hydrolase	<b>Query Match</b> Best Local Similarity      39.6%; Score 634.5; DB 2; Length 219; 54.8%; Pred. No 5.4e-47;
---	--





Query Match 36.9%; Score 591; DB 2; Length 256;  
 Best Local Similarity 53.2%; Pred. No. 3.6e-43;  
 Matches 116; Conservative 31; Mismatches 63; Indels 8; Gaps 3;

```

QY 85 TWREELAAEFKEPKYKQMSFVADERSRH---HYVPADQVYSTMCIDIDYKVYILG 140
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 24 SMKAVHGNMILLRPMRLDLSAFL--RAKVAAGSVYPPGSGIYFAEFATPPQRYKAVILG 80
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 141 ODYPHGNQAHGLCFSVQAKVPPPPSLVNIYKELCTDIDGFKHPGHDLSGMAKQVLLL 200
   ||||| ||||| ||||| ::|||::|||::|::|::|::|::|::|::|::|::|
Db 81 QDPYHGGQAHGLCFSVRPGWPLRPSLNTYKLEEDL-GLLRPDHGCLLPWAKRGVLLL 139
   ||||| ::|||::|||::|::|::|::|::|::|::|::|::|::|::|::|
QY 201 NAVITVAHQANSKDKGMEFTDPAVIKWL SVNREGVFLMGSYAHKKATIDRRHHV 260
   ||||| ::|||::|||::|::|::|::|::|::|::|::|::|::|::|::|
Db 140 NAVITVEDGRAGAHQGWGEGFTDHYVDTLNRRERGLVFMIMGSYAQAKGVIDRRHLV 199
   ||||| ::|||::|||::|::|::|::|::|::|::|::|::|::|::|::|
QY 261 LOAVHPSPLSAHRGFLGCKHPSKANGLLKLGTEPTNW 298
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 200 LKAHPSPLSAHRGFLGCKRHPSLCNOYLSOHGLGMVDM 237
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

## RESULT 15

uracil-DNA glycosylase PA0750 [imported] - Pseudomonas aeruginosa (strain PA01)  
 C83551  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: C83551  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B.  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
 .; Lory, S.; Olson, M.V  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
 A:Reference number: A82950; MUID:20437337  
 A:Accession: C83551  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-231 <STO>  
 A:Cross-references: GB:AE004510; GB:AE004091; NID:9946634; PIDN:AAG04139.1; GSPDB:GN001  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: ung; PA0750  
 C:Superfamily: uracil-DNA glycosylase

Query Match 36.6%; Score 586.5; DB 2; Length 231;  
 Best Local Similarity 49.8%; Pred. No. 7.6e-43;  
 Matches 114; Conservative 40; Mismatches 66; Indels 9; Gaps 3;

```

QY 71 DKIRAKATPAGFETWRRELAAEFEKPYFKOLMSFVADERSR-HTVYPPADQVYSTMC 129
   ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 6 DRIKLEA-----SMKEALREEFEDKPYMKOLGEFLROEKAKGVIFPPGPLIFNALNTT 58
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 130 DIDQVYVITIGDPPYHGNQAHGLCFSVQAKVPPPPSLVNIYKELCTDIDGFKHPGHDL 189
   ::|||::|||::|::|::|::|::|::|::|::|::|::|::|::|
Db 59 PLENVAVYITIGDPPYHGNQAHGLCFSVQAKVPPPPSLVNIYKELNRLDN-IPIPNNGYL 117
   ::|||::|||::|::|::|::|::|::|::|::|::|::|::|::|
QY 190 SGWAKQGVLLINAVLYVRAHQANSKDKGMEFTDPAVIKWL SVNREGVFLMGSYAHKK 249
   ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 118 QRWAEQGVLLINTSLTVEQAKAGSHANAGWQPTDRIEYVNERCERLVFLMGSHAQSK 177
   ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 250 GATIDRRKHVLAQVHPSPLSAHRGFLGCKHPSKANGLLKLGTEPTNW 298
   ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 178 QKLIDPQRHILTKSAHPSPLSAHRGFLGNGHFSRTNKFLEONGKPTIDW 226
   ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

Search completed: August 21, 2002, 10:25:17  
 Job time: 184 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 21, 2002, 10:24:58 ; Search time 13.44 Seconds

(without alignments)  
867.157 Million cell updates/sec

Title: US-09-758-017a-2

Perfect score: 1603

Sequence: 1 MFLKGLGCRICISSNRVLPG.....SKANGLLKLGTEPINMRAL 301

Scoring table: BLOSUM62

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database: SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1051.5	65.6	295	1 UNG_MOUSE	P97931 mus musculu
2	1032	64.4	304	1 UNG_HUMAN	P13051 homo sapien
3	681.5	42.5	228	1 UNG_ECOLI	P12295 escherichia
4	665.5	41.5	222	1 UNG_PASMU	P57807 pasteurilla
5	635.5	39.6	219	1 UNG_HAETN	P43731 haemophilus
6	601	37.5	257	1 UNG_DICDI	P33766 dictyosteli
7	597	37.2	217	1 UNG_PSEDE	P29950 pseudomonas
8	572.5	35.7	224	1 UNG_BACHD	O9K682 bacillus ha
9	570.5	35.6	225	1 UNG_BACRO	P19615 bacillus su
10	561	35.0	322	1 UNG_SCHRO	O74834 schizosacch
11	552	34.4	229	1 UNG_CHLMU	O9P342 chlamydia m
12	551	34.4	220	1 UNG_BUCAI	P57280 buchera ap
13	548	34.2	229	1 UNG_CHLTR	O84613 chlamydia t
14	540	33.7	236	1 UNG_YEAST	P12887 saccharomyc
15	539.5	33.7	236	1 UNG_CHLPN	O92743 chlamydia p
16	537.5	33.5	217	1 UNG_STRAG	O9X468 streptococc
17	535.5	33.4	217	1 UNG_STRPN	P23379 streptococc
18	496.5	31.0	223	1 UNG_BORBU	O51082 borrelia bu
19	472.5	29.9	312	1 UNG_HSVB	P28866 equine herp
20	463.5	28.9	255	1 UNG_EBV	P12888 Epstein-Bar
21	457	28.5	255	1 UNG_HSV2	P53755 equine herp
22	454.5	28.4	219	1 UNG_LACIA	O9C182 lactococcus
23	454.5	28.4	315	1 UNG_PRTVE	P22506 pseudorabie
24	454.5	28.4	316	1 UNG_PRYKA	P22507 pseudorabie
25	445	27.8	252	1 UNG_HSV4	O01019 herpesvirus
26	429.5	26.8	250	1 UNG_HCMVA	P10166 human cytom
27	423	26.4	334	1 UNG_HSV1	P10166 herpes simp
28	422	26.3	255	1 UNG_HSV2H	P28275 herpes simp
29	415	25.9	235	1 UNG_HSV6U	P23245 human herpe
30	411.5	25.7	204	1 UNG_HSVBC	P23245 bovine herp
31	411	25.6	255	1 UNG_HSV62	P22447 human herpe
32	408.5	25.5	254	1 UNG_HSV7J	P50639 human herpe
33	401	25.0	294	1 UNG_HSV23	P13158 herpes simp

34	396	24.7	233	1 UNG_HELPU	O9ZJ99 helicobacte
35	388.5	24.2	305	1 UNG_VZVD	P09307 varicella-z
36	384	24.0	233	1 UNG_HELPU	P56397 helicobacte
37	367	22.9	227	1 UNG_MYCRU	P51119 mycobacteri
38	359	22.4	227	1 UNG_MYCRU	O9C833 mycobacteri
39	247.5	15.4	240	1 UNG_MYCPN	P75536 mycoplasma
40	244.5	15.3	245	1 UNG_MYCGE	P47343 mycoplasma
41	97.5	6.1	218	1 UNG_FOWPV	P21968 fowlpox vlr
42	94	5.9	721	1 ERG7_SCHRO	O10231 schizosacch
43	92.5	5.8	836	1 Y167_HUMAN	O99490 homo sapien
44	92	5.7	577	1 MTGB_MOUSE	O61909 mus musculu
45	89	5.6	218	1 UNG_VACCV	P04303 vaccinia vi

## ALIGNMENTS

```

RESULT 1
UNG_MOUSE STANDARD: PRT: 295 AA.
AC P97931; P97509;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Utracil-DNA glycosylase, mitochondrial precursor (EC 3.2.2.-) (UDG).
GN UNG OR UNG1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97311407; PubMed=9168124;
RA Svendsen P.C., Yee H.A., Winkfeld R.J., van de Sande J.H.;
RT "The mouse uracil-DNA glycosylase gene: isolation of cDNA and genomic
RL clones and mapping ung to mouse chromosome 5.";
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97169285; PubMed=9016624;
RA Nilsen H., Solum K., Haug T., Krokan H.E.;
RT "Nuclear and mitochondrial uracil-DNA glycosylases are generated by
RL alternative splicing and transcription from different positions in
RN the ung gene.";
RX Nucleic Acids Res. 25:750-755(1997).
CC -1- FUNCTION: EXCISES URACIL RESIDUES FROM THE DNA WHICH CAN ARISE
CC AS A RESULT OF MISINCORPORATION OF DUMP RESIDUES BY DNA
CC POLYMERASE OR DUE TO DEAMINATION OF CYTOSINE.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE URACIL-DNA GLYCOSYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U55040; AB39511.1; -
DR EMBL: U55041; AAC53197.1; -
DR EMBL: X99018; CA67489.1; -
DR HSSP: P13051; IAKZ.
DR MGD: MGI:109352; Ung.
DR InterPro: IPR000243; U_DNA_glycsylase.
DR InterPro: IPR0003249; Utrac_DNA_glycsyl.
DR Pfam: PF00315; UNG. 1.
DR PRODOM: PD001589; Utrac_DNA_glycsyl. 1.
DR PROSITE: PS00130; U_DNA_GLYCOSYLASE. 1.
KW DNA repair; Hydrolase; Glycosidase; Nuclear protein; Mitochondrion;
KW Transil peptide.
FT TRANSIT 1 58 MITOCHONDRION (POTENTIAL).

```





DR EMBL: D64044; BAA10923.1; -  
 DR EMBL: AE000344; AAC75633.1; -  
 DR EMBL: D90886; BAA16466.1; -  
 DR PIR: A28175; DSECU  
 DR PDB: 1EUG; 12-OCT-99.  
 DR PDB: 2EUG; 13-OCT-99.  
 DR PDB: 3EUG; 13-OCT-99.  
 DR PDB: 4EUG; 23-JUL-99.  
 DR PDB: 5EUG; 23-JUL-99.  
 DR PDB: 1EUI; 22-JUN-99.  
 DR PDB: 1UUG; 25-MAR-99.  
 DR PDB: 2UUG; 25-MAR-99.  
 DR PDB: 1FLZ; 17-JAN-01.  
 DR Ecogen; Egl1058; ung.  
 DR InterPro: IPR002043; U\_DNA\_glycosylase.  
 DR InterPro: IPR003249; Urac\_DNA\_glycosyl.  
 DR Pfam: PF00315; UNG; 1.  
 DR ProDom; PD001589; Urac\_DNA\_glycosyl; 1.  
 DR PROSITE; PS00130; U\_DNA\_GLYCOSYLASE; 1.  
 KW DNA repair; Hydrolase; Glycosidase; 3D-structure; Complete proteome.  
 FT INIT\_MET 0  
 FT ACT\_SITE 63 63  
 SQ SEQUENCE 228 AA; 25562 MW; E03126BF78F44BB1 CRC64;

Query Match 42.5%; Score 681.5; DB 1; Length 228;  
 Best Local Similarity 60.9%; Pred. No. 2.3e-53;  
 Matches 131; Conservative 24; Mismatches 59; Indels 1; Gaps 1;

QY 85 TWKRELAAEFKRYFKQLMSFVADER-SRRTVYPPADQVYSTMCIDIDYKVYVILGOD 143  
 Db 5 TWKDVLAEEKQDPYFELTQTVASERQSVTIYPPQKDVNAFRETGLGDKVYVILGOD 64  
 QY 144 YHGPNOAHGICFSYQKRPVPPPSLVNRYKELCTDIDGFKHFGHDLGMAKOGVLLNAV 203  
 Db 65 YHGPNOAHGIAFSYRPGIAPRPSLNMRYKLENTIPGTRNHYLSMARQGVLLNTV 124  
 QY 204 LTVRAHQASHKDRGWFETFDVAIKWLSVNBEGVFLMGSYAHKKGATIDRRKHVLA 263  
 Db 125 LTVRAGQASHASISGWFETFDKYSILINOREGVFLMGSHAKGAIIDKORHVLKA 184  
 QY 264 VHPSPLSAHRGFLGCRHFSKANGLLKSGTEPINW 298  
 Db 185 PHPSPLSAHRGFGCNHFVLANOMLEORGETPIDW 219

RESULT 4  
 UNG\_PASMU  
 ID UNG\_PASMU STANDARD; PRT; 222 AA.  
 AC P57807;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, last sequence update)  
 DT 16-OCT-2001 (Rel. 40, last annotation update)  
 DE Uracil-DNA glycosylase (EC 3.2.2.-) (UDG).  
 GN UNG OR PM0065  
 OS Pasteurella multocida.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Pasteurella.  
 OX NCBI\_TaxID=747;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PM70;  
 RX MEDLINE=21145866; PubMed=11248100;  
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;  
 RT "Complete genomic sequence of Pasteurella multocida PM70.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
 CC -1- FUNCTION: EXCISES URACIL RESIDUES FROM THE DNA WHICH CAN ARISE AS  
 CC A RESULT OF MISINCORPORATION OF DUMP RESIDUES BY DNA POLYMERASE OR  
 CC DUE TO DEAMINATION OF CYTOSINE (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE URACIL-DNA GLYCOSYLASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: AE006042; AAK02149.1; -  
 DR InterPro: IPR002043; U\_DNA\_glycosylase.  
 DR InterPro: IPR003249; Urac\_DNA\_glycosyl.  
 DR Pfam: PF00315; UNG; 1.  
 DR ProDom; PD001589; Urac\_DNA\_glycosyl; 1.  
 DR PROSITE; PS00130; U\_DNA\_GLYCOSYLASE; 1.  
 KW DNA repair; Hydrolase; Glycosidase; Complete proteome.  
 FT ACT\_SITE 61 61  
 SQ SEQUENCE 222 AA; 24906 MW; 3F9320277FE33367 CRC64;

Query Match 41.5%; Score 665.5; DB 1; Length 222;  
 Best Local Similarity 58.1%; Pred. No. 5.9e-52;  
 Matches 126; Conservative 29; Mismatches 61; Indels 1; Gaps 1;

QY 84 ETWRELAAEFKRYFKQLMSFVADER-SRRTVYPPADQVYSTMCIDIDYKVYVILGOD 142  
 Db 2 KTWKDVIGTEKTPQYFKHILIDVQARASGRIVYPPQEVESARQLEFEVAKVYVILGOD 61  
 QY 143 YHGPNOAHGICFSYQKRPVPPPSLVNRYKELCTDIDGFKHFGHDLGMAKOGVLLNA 202  
 Db 62 YHGPNOAHGIAFSYRPGIAPRPSLNMRYKLENTIPGTRNHYLSMARQGVLLNT 121  
 QY 203 LTVRAHQASHKDRGWFETFDVAIKWLSVNBEGVFLMGSYAHKKGATIDRRKHVLA 262  
 Db 122 VLVTEQCAKASHASISGWFETFDRIYALINOREKLVFLMGSHAKGQFIDRKHCVET 181  
 QY 263 AVHPSPLSAHRGFLGCRHFSKANGLLKSGTEPINW 299  
 Db 182 APHPSPLSAHRGFLGCRHFSKNTAYLMAQGLSPLOW 218

RESULT 5  
 UNG\_HAEIN  
 ID UNG\_HAEIN STANDARD; PRT; 219 AA.  
 AC P43731;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, last sequence update)  
 DT 16-OCT-2001 (Rel. 40, last annotation update)  
 DE Uracil-DNA glycosylase (EC 3.2.2.-) (UDG).  
 GN UNG OR HT0018  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RD / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Gload A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Ucterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Smill K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus  
 RT influenzae Rd.";  
 RL Science 269:496-512(1995).  
 CC -1- FUNCTION: EXCISES URACIL RESIDUES FROM THE DNA WHICH CAN ARISE  
 CC AS A RESULT OF MISINCORPORATION OF DUMP RESIDUES BY DNA  
 CC POLYMERASE OR DUE TO DEAMINATION OF CYTOSINE (BY SIMILARITY).  
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (BY similarity).  
 CC -----

```
CC -1- SIMILARITY: BELONGS TO THE URACIL-DNA GLYCOSYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32687; AAC21696.1; -.
CC HSSP: P12295; 3EUG.
CC TIGR: HT0018; -.
CC InterPro: IPR002043; U-DNA_glycosylase.
CC InterPro: IPR003249; Urac_DNA_glycosyl.
CC Pfam: PF00315; UNG; 1.
CC ProDom: PD001589; Urac_DNA_glycosyl; 1.
CC PROSITE: PS00130; U-DNA_GLYCOSYLASE; 1.
CC DNA_repair; Hydrolase; Glycosidase; Complete proteome.
CC ACT_SITE 61 61 GENERAL BASE (BY SIMILARITY).
CC SEQUENCE 219 AA; 24848 MW; C270CC31F7B2E58A CRC64;

Query Match
Best Local Similarity 39.6%; Score 635.5; DB 1; Length 219;
Matches 124; Conservative 28; Mismatches 64; Indels 1; Gaps 1;

QY 84 ETRRELAAEFEPYFKOLMSFVADER-SRHTVYPADQVYSTEMCDIODVKVILGOD 142
DB 2 KMTDVIIGEKAPRYQHTLQGVHLARASGKTIYRPQEDVFNAPKTAFFDVAVVILGOD 61
QY 143 PYHGPNQAHGICFSYOKRPVPPPSLVNITKELCTDIDGFKHGHGDISGNAKGVLLNA 202
DB 62 PYHGPNQAHGICFSYOKRPVPPPSLVNITKELCTDIDGFKHGHGDISGNAKGVLLNA 121
QY 203 VLTVRANQAHNSHDKRGMEPTDAVIKMLSVNREGVFLMGSAHKKGATIDRRHHVLQ 262
DB 122 VLTVERGMASHANLGMERTDKVIAVLNHRKLVFLLMGSHAOKKGOMIDRRLHVL 181
QY 263 AVHPSPLSAHRGFLGCKHFSKANGLKLSGTEPINMR 299
DB 182 APHPSPLSAHRGFGFCGRHFSKTSYLSHGIRPIDWQ 218

RESULT 6
UNG_DICDI
ID UNG_DICDI STANDARD; PRT; 257 AA.
AC P53766;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Uracil-DNA glycosylase (EC 3.2.2.-) (UDG) (Fragment).
GN UGLA.
OS Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OC NCBI_Taxid=44689;
RX NCBI_Taxid=44689;
RN [1]
RP SEQUENCE FROM N.A.
RA Guyer R.B., Deering R.A.;
RL Submitted (SEP-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: EXCISES URACIL RESIDUES FROM THE DNA WHICH CAN ARISE
CC AS A RESULT OF MISINCORPORATION OF DUMP RESIDUES BY DNA
CC POLYMERASE OR DUE TO DEAMINATION OF CYTOSINE.
CC -1- SIMILARITY: BELONGS TO THE URACIL-DNA GLYCOSYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
DR EMBL: U32866; AAA75334.1; -.
DR HSSP: P13051; 1AKZ.
DR DicyDb: DD05081; uglA.
DR InterPro: IPR002043; U-DNA_glycosylase.
DR InterPro: IPR003249; Urac_DNA_glycosyl.
DR Pfam: PF00315; UNG; 1.
DR ProDom: PD001589; Urac_DNA_glycosyl; 1.
DR PROSITE: PS00130; U-DNA_GLYCOSYLASE; 1.
DR DNA_repair; Hydrolase; Glycosidase; DNA repair.
CC ACT_SITE 1 1 POLY-ASN.
CC ACT_SITE 97 97 GENERAL BASE (BY SIMILARITY).
CC SEQUENCE 257 AA; 29516 MW; 9CFA0898B648759 CRC64;

Query Match
Best Local Similarity 37.5%; Score 601; DB 1; Length 257;
Matches 117; Conservative 33; Mismatches 63; Indels 4; Gaps 3;

QY 86 WRRELAAEFKPYFKOLMSFVADERS--RHHTVYPADQVYSTEMCDIODVKVILGODP 143
DB 39 WKRELSGEFGKAYFKKMITOLNKRYSSEKRPYPPKNEIFSAYVAHLEDVKVILGODP 98
QY 144 YHGPNQAHGICFSYOKRPVPPPSLVNITKELCTDIDGFKHP-GHGDLSGNAKGVLLNA 202
DB 99 YHGQGNHGLSFSYKKGVSPPSLNITKELTDIEGFKPLKGLFEPWARGGVLLNA 158
QY 203 VLTVRANQAHNSHDKRGMEPTDAVIKMLSVNREGVFLMGSAHKKGATIDRRHHVLQ 262
DB 159 VLTVEEATPSHNDGFAADTDAVLKLSKQDQPIVILMGFGAKKKEFTNNHVLK 218
QY 263 AVHPSPLSAHRGFLGCKHFSKANGLKLSGTEPINMR 299
DB 219 SGHPSPLSI-KHFGCKHFSKNEFLKSGIEIDWK 254

RESULT 7
UNG_PSEDE
ID UNG_PSEDE STANDARD; PRT; 217 AA.
AC P29950;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Uracil-DNA glycosylase (EC 3.2.2.-) (UDG) (Fragment).
GN UNG.
OS Pseudomonas denitrificans.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_Taxid=43306;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=SC510;
RX MEDLINE=91008976; pubmed=2211521;
RA Crouzet J., Cameron B., Cauchois L., Rigault S., Rouyez M.-C.,
RA Blanche F., Thibaut D., Debussche L.;
RT "Genetic and sequence analysis of an 8.7-kilobase Pseudomonas
RT denitrificans fragment carrying eight genes involved in
RT transformation of precorrin-2 to cobyrinic acid.";
RL J. Bacteriol. 172:5980-5990(1990).
RN [2]
RP IDENTIFICATION, AND CORRECTION OF A PROBABLE FRAMESHIFT.
RA Asland R.;
RL Unpublished observations (FEB-1993).
CC -1- FUNCTION: EXCISES URACIL RESIDUES FROM THE DNA WHICH CAN ARISE
CC AS A RESULT OF MISINCORPORATION OF DUMP RESIDUES BY DNA
CC POLYMERASE OR DUE TO DEAMINATION OF CYTOSINE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE URACIL-DNA GLYCOSYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

	modified and this statement is not removed.	Usage by	and for commercial entitles requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb.sib.ch).
CC			
CC			
CC			
CC			
CC			
DR	EMBL; M59301;	-;	NOT_ANNOTATED_CDS.
DR	PIR; I36145;	I36145.	
DR	HSSP; P12295;	1EUQ.	
DR	InterPro; IPRO02043;	U_DNA_glycylase.	
DR	InterPro; IPRO03249;	Urac_DNA_glycosyl.	
DR	pfam; PF00315:	UNG; 1.	
DR	ProdDom; PD001589;	Urac_DNA_glycosyl; 1.	
DR	PROSITE; PS00130;	U_DNA_GLYCOSYLASE; 1.	
RW	DNA_repair:	Hydrolase; Glycosidase.	
KW	NON_TER	1	
FT	ACT_SITE	51 .. 51	GENERAL BASE (BY SIMILARITY).
SQ	SEQUENCE	217 AA; 24314 MW; CCGCB5491861AC4	CNC64;
	Query Match	37.2%; Score 597;	DB 1; Length 217;
	Best Local Similarity	55.3%;	Pred. No. 6.9e-46;
	Matches 115;	Conservative 28;	Mismatches 63; Indels 2; Gaps 2;
OY	EPEKPYFQLNSF-VADESRHTTVPPADVOYSSTSEMCDIDQKVILLGDDPHGNPAH	151	
Db	:  :  :	:	:  :
	1 EFASAYMADLKOFVLAQNKGHRQTTPRCPETFRALDTLPDKRVYLLIGDPYHGDAQH	60	
OY	GLCFSVQRVPPLPSLVNYIKELCTDIDGFRRPGHDLSGNAQGVLNLNAVLTVAHQO	211	
Db	:              :	:	:        :  :
	61 GLCFSVRGVRPTPLSLNVIKELMTDL-GIPPARHFLESNARGCVLTLNSLVTRARER	119	
OY	NSHKRGMETFLDAIYIKLSVNREGVVALLGSYNARKGATITDRKRHHVALDVHPSPISA	271	
Db	:: ::           :  :	:	::
	120 ASHQGHGEKEKTDFAIIRVANAEHVPMWLGSTAQKKAAFLVDNRSHLTRAPHSPLSA	179	
OY	272 HRFGLCKRHFSKANGLKLSTGPENMR	239	
Db	:      :  :  :		
	180 HSGFLGRHFESQANAFLESKGFDPDMR	207	
RESULT	8		
ID	UNG_BACHD	STANDARD;	PRT; 224 AA.
AC	O9K682;		
DC	16-OCT-2001 (Rel. 40, Created)		
DE	16-OCT-2001 (Rel. 40, Last sequence update)		
DI	16-OCT-2001 (Rel. 40, Last annotation update)		
DN	Uracil-DNA glycosylase (EC 3.2.2.-) (UDS).		
GZ	UNC OR BH3850.		
OS	Bacillus halodurans.		
OC	Bacteria; Firmicutes; Bacilllus/Clostridium group;		
CC	Bacillus/Staphylococcus group; Bacilllus.		
OX	NCBI_TaxId=86665;		
RM	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C-125 / JCM 9153;		
RX	MEDLINE=20512582; PubMed=11058132;		
RA	Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,		
RA	Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,		
RA	Horiikoshi K.;		
RT	"Complete genome sequence of the alkaliphilic bacterium Bacillus		
RT	halodurans and genomic sequence comparison with Bacillus subtilis."		
RL	Nucleic Acids Res. 28:4317-4311(2000).		
CL	-1- FUNCTION: EXCISES URACIL RESIDUES FROM THE DNA WHICH CAN ARISE		
CC	AS A RESULT OF MISINCORPORATION OF DUMP RESIDUES BY DNA		
CC	POLYMERASE OR DUE TO DEAMINATION OF CYTOSINE (BY SIMILARITY).		
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).		
CC	-1- SIMILARITY: BELONGS TO THE URACIL-DNA GLYCOSYLASE FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
CC	The European Bioinformatics institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to: license@sib-sib.ch).
CC -----
DR   EMBL; AP001520; BAB07569.1; -
DR   InterPro; IPR002043; U_DNA_glycylase.
DR   InterPro; IPR003249; Uraci_DNA_glycosyl.
DR   Pfam; PF00315; UNG; 1.
DR   ProDom; PS001589; Uraci_DNA_glycosyl; 1.
DR   PROSITE; PS00130; U_DNA_GLYCOSYLASE; 1.
KW   DNA repair; Hydrolase; Glycosidase; Complete proteome.
FT   ACT_SITE      64
SQ   SEQUENCE      224 AA: 25582 MW; DF8A096F1235E605 CRC64;

Query Match          35.7%; Score 572.5; DB 1; Length 224;
Best Local Similarity 52.1%; Pred. No. 1,le-43;
Matches 111; Conservative 32; Mismatches 69; Indels 1; Gaps 1;

QY    86 WRRELAEEFEKPYFKOLMSFNADERSRHTVYPAPDQVYSTEMCDIODVKVILGDDPYH 145
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB    8 WADVLEAEFEOKRYVQLREFELKOEYSTOTIYPSMDIFALHYTPLAARAKVILLGDPPH 67
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY    146 GPNOAHGICFSYOKRVPPRPSPVNYIKELCTNDIGCFKHGHGGDSGMWKGCVLLNAVLT 205
       ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB    68 GPNOAHGHSFSYKPGVPRLPSPLONTLFKELRDL-CGPLPNNHGLTLPWDQGLLVNLYLT 126
       ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY    206 VRAHQASHKRDGWETFDVAVIKMLSVNRGEVFPFLMGSYAAHKCATIDRRKHHVLQAVH 265
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB    127 VKRGQAASHRGKGMSETFDVRITCLNEREKPIYFIWMGRHAQEKQALITNDRHYILTAPH 186
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY    266 PSPPLAHNGFLGCKHFPSKANGLKLKSCTEPINW 298
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB    187 PSPFSANRGFEGRPFSTINRILOEQEGESQHW 219
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT
9
UNG_BACSU
AC   ID            UNG_BACSU           STANDARD:         PRT:        225 AA.
AC   P33615.
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Uracil-DNA glycosylase (EC 3.2.2.-) (UDC).
GN   URC OR IPA-57D.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168.
RX   MEDLINE=95020537; PubMed-7934828.
RA   Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,
RA   Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
RA   Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
RA   Rapoport G., Danchin A.;
RT   "Bacillus subtilis genome project: cloning and sequencing of the 97
RT   kb region from 325 degrees to 333 degrees.";
RL   Mol. Microbiol. 10:371-384(1993).
CC   -!- FUNCTION: EXCISES URACIL RESIDUES FROM THE DNA WHICH CAN ARISE
CC   AS A RESULT OF MISINCORPORATION OF DUMP RESIDUES BY DNA
CC   POLYMERASE OR DUE TO DEAMINATION OF CYTOSINE (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE URACIL-DNA GLYCOSYLASE FAMILY.
CC -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to: license@sib-sib.ch).
```









```

CC      AS A RESULT OF MISINCORPORATION OF DUMP RESIDUES BY DNA
CC      POLYMERASE OR DUE TO DEAMINATION OF CYTOSINE (BY SIMILARITY).
CC      -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO THE URACIL-DNA GLYCOSYLASE FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AE001659; AAD18911.1; -
DR      EMBL; AE002266; AAF38868.1; -
DR      EMBL; AP002547; BAA98981.1; -
DR      HSSP; P12295; 1EUG.
DR      TIGR; CP1099; -
DR      InterPro; IPR002043; U_DNA_glycosylase.
DR      InterPro; IPR003249; Urac_DNA_glycosyl.
DR      Pfam; PF00315; UNG; 1.
DR      ProDom; PD001589; Urac_DNA_glycosyl; 1.
DR      PROSITE; PS00130; U_DNA_GLYCOSYLASE; 1.
KM      DNA repair; Hydroxylase; Glycosidase; Complete proteome.
FT      ACT_SITE 77
FT      ACT_SITE 77
SQ      SEQUENCE 236 AA; 27006 MW; 1A734AD68B81AA03 CRC64;

```

```

Query Match      33.7%; Score 539.5; DB 1; Length 236;
Best Local Similarity 46.2%; Pred. No. 9.9e-41;
Matches 111; Conservative 34; Mismatches 82; Indels 13; Gaps 4;

```

```

QY      66 KKAALDKIRAKAPPAGE---TWRELAEEFEKPYFKQLMSFVADERSRHTVYPPADQ 121
DB      2 QNATIDQL-----FVSWQEQPLPLCWREQLKEEMSKPYMQQLIFLKQEKHEHTVYPEENC 56
QY      122 VYSTEMCDIODYKVVITLGDPRHGPNQAHGCLCSVOKPYPPPSLVNITKELCTDIDGF 181
DB      57 VFSALNRSTPPDYRVVITLGDPPYPGKGAHGLSFSVPEGORLPPSLINIFRELKTDLGIE 116
QY      182 KHPGHGDLGSGMAKGVLLNAVLTVRAHQANSHKDRGMETFTDAVIKWLNVNREGVVEL 241
DB      117 NH--KGCLOGMANOGIILLNTVLTVRAGEPFPSHAGKGMELFTDAIVTKLIQERHITFVL 174
QY      242 WGSYAHKKGAFI--DRKRHHVLAQVHPSPLSAHRGFLGCKHFSKANGLLKLGSTEPINMR 299
DB      175 WGAARAKKCELLFNRSKHQHAVLSSPHSPPLAHRGFGCSHFSKINYLNLKLNKPMINMK 234

```

Search completed: August 21, 2002, 10:28:51  
 Job time: 233 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 21, 2002, 10:24:38 ; Search time 30.24 Seconds

(without alignments)  
1721.940 Million cell updates/sec

Title: US-09-758-017a-2

Perfect score: 1603  
Sequence: 1 MLEKLGICQRCISSNRVLPG.....SKANGLLKLGSTEPINMRAL 301

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SP:Archea:\*
- 2: SP:Bacteria:\*
- 3: SP:Fungi:\*
- 4: SP:Human:\*
- 5: SP:Invertebrate:\*
- 6: SP:Mammal:\*
- 7: SP:Mhc:\*
- 8: SP:Organelle:\*
- 9: SP:Phage:\*
- 10: SP:Plant:\*
- 11: SP:Rodent:\*
- 12: SP:Virus:\*
- 13: SP:Vertebrate:\*
- 14: SP:Unclassified:\*
- 15: SP:Virus:\*
- 16: SP:Bacteriaph:\*
- 17: SP:Archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1596	99.6	301	13	Q91983 gadus morhu
2	1421	88.6	301	13	Q91982 gadus morhu
3	1051	65.6	306	11	Q91TW8 mus musculu
4	1045	65.2	306	11	P97285 mus musculu
5	1018.5	65.3	313	4	Q93028 homo sapien
6	887	55.3	199	11	Q99KR9 mus musculu
7	702.5	43.8	226	16	Q9KPK8 vibrio chol
8	648.5	40.5	219	16	Q9JZAI neisseria m
9	635.5	39.6	247	16	Q9RMH9 Ogrwh9 deinococcus
10	634.5	39.6	219	16	Q9JUC4 neisseria m
11	632.5	39.5	282	5	Q9U221 caenorhabdi
12	606	37.8	241	16	Q921U5 rhizobium m
13	598.5	37.3	330	10	Q9LTH6 arabidopsis
14	591	36.9	227	2	Q9K320 streptomyce
15	591	36.9	256	16	Q9PA28 xyella fas
16	586.5	36.6	231	16	Q915H9 pseudomonas

17	569	35.5	221	16	Q9PU40 campylobact
18	559.5	34.9	218	16	Q92C11 listeria in
19	557	34.7	218	16	Q99W30 staphylococ
20	555.5	34.7	298	5	Q9U776 trypanosoma
21	536.5	33.5	221	16	Q92E00 listeria in
22	528.5	33.0	217	16	Q9A072 streptococc
23	474	29.6	314	12	Q93301 equine herp
24	472	29.4	252	12	Q93695 atelaphine
25	467.5	29.2	256	12	Q993H3 callitrichi
26	466.5	29.1	254	12	Q69273 leporid her
27	463	28.9	253	12	Q99C24 bovine herp
28	458.5	28.6	255	12	Q80892 herpesvirus
29	451	28.1	252	12	Q9YTL9 ateline her
30	446.5	27.9	212	16	Q9PPU2 ureaplasma
31	445.5	27.8	301	12	Q93497 bovine herp
32	445	27.7	273	12	Q91T17 tupala herp
33	444.5	27.7	255	12	Q9WR05 macaca mula
34	436	27.2	249	12	P88934 kaposi's sa
35	435.5	27.2	249	12	P88984 murid herpe
36	423	26.4	244	12	Q9YFJ3 human herpe
37	422.5	26.1	300	12	Q9WR44 cercoplithec
38	418.5	26.1	259	12	Q9DM85 rat cytomec
39	418.5	26.1	323	12	Q9QTE2 gallid herp
40	415	25.9	293	12	Q37930 feline herp
41	409	25.5	255	12	Q69070 human herpe
42	406	25.3	255	12	Q57142 human herpe
43	402.5	25.1	221	16	Q98PV4 mycoplasma
44	389	24.3	305	12	Q9J3N2 human herpe
45	384.5	24.0	313	12	Q9E6R2 turkey herp

ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	301 AA.
Q91983	Q91983			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	URACIL-DNA GLYCOSYLASE (EC 3.2.2.3).			
GN	UNG1.			
OS	Gadus morhua (Atlantic cod).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;			
OC	Acanthomorpha; Paracanthopterygii; Gadiformes; Gadoidae; Gadidae;			
OC	Gadus.			
OX	NCBI_TaxID=8049;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=LIVER.			
RA	Lanes O., Willassen N.P.;			
RT	"Identification, cloning and expression of nuclear and mitochondrial			
RT	uracil-DNA glycosylase from Atlantic cod(Gadus morhua).			
RT	Characterisation of the cold-active catalytic domain.";			
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AJ275971; CAB85707.1; -.			
DR	HSSP: P13051; 1AK2.			
DR	InterPro: IPR003249; Urac_DNA_glycosyl.			
DR	InterPro: IPR002043; U_DNA_glycosylase.			
DR	Pfam: PF00315; UNG; 1.			
DR	ProDom: PD001589; Urac_DNA_glycosyl; 1.			
DR	PROSITE: PS00130; U_DNA_GLYCOSYLASE; 1.			
KW	Hydrolase; Glycosidase.			
SQ	SEQUENCE 301 AA; 33807 MW; 21866ADNAAD96C5 CRC64;			

Query Match 99.6%; Score 1596; DB 13; Length 301;  
Best Local Similarity 99.7%; Pred. No. 1.1e-142;  
Matches 300; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Db 1 MRLKLGICORCISNRNLPGLLPQTLCFSELMKITPKLRSSVNEOKTSSPOLSVPOLE 60
QY 61 RMAKNKKAALDKIRAKATTPAGFETWRRELAEPKPYFKQOLMSFVADERSRHTVYPAD 120
Db 61 RMAKNKKAALDKIRAKATTPAGFETWRRELAEPKPYFKQOLMSFVADERSRHTVYPAD 120
QY 121 QVYSSTEMCDIODYKVYVILGODPYHGNQAHGLCFSVQKRPVPPPSLVNITKELCTDIDG 180
Db 121 QVYSSTEMCDIODYKVYVILGODPYHGNQAHGLCFSVQKRPVPPPSLVNITKELCTDIDG 180
QY 181 FKPHGHGDLDSGMAKQGVLLNNAVITVRAHOANSKDKGWEFTDAVTKWLSVNEGVFL 240
Db 181 FKPHGHGDLDSGMAKQGVLLNNAVITVRAHOANSKDKGWEFTDAVTKWLSVNEGVFL 240
QY 241 LMGSYAAKKGATIDRKRNHVLQAVHPSPLSAHRGFLGCKHFSKANGLLKLSGTEPIWMRA 300
Db 241 LMGSYAAKKGATIDRKRNHVLQAVHPSPLSAHRGFLGCKHFSKANGLLKLSGTEPIWMRA 300
QY 301 L 301
Db 301 L 301
```

```
RESULT 2
Q91982 PRELIMINARY; PRT; 301 AA.
ID Q91982;
AC Q91982;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE URACIL-DNA GLYCOSYLASE (EC 3.2.2.3).
GN UNG2.
OS Gadus morhua (Atlantic cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae;
OC Gadus.
OX NCBI_TaxID=8049;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Lanes O., Willassen N.P.;
RT "Identification, cloning and expression of nuclear and mitochondrial
RT uracil-DNA glycosylase from Atlantic cod(Gadus morhua).";
RT Characterisation of the cold-active catalytic domain.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ275972; CAB85708.1; -.
DR HSSP; P13051; IAKZ.
DR InterPro: IPR003249; Urac-DNA_glycosyl.
DR InterPro: IPR002043; U_DNA_glycosylase.
DR Pfam; PF00315; UNG; 1.
DR ProDom; PD001589; Urac_DNA_glycosyl; 1.
DR ProSITE; PS00130; U_DNA_GLYCOSYLASE; 1.
KW Hydroxylase; Glycosidase.
SQ SEQUENCE 301 AA; 33853 MW; C5F0BBALDSCA89B CRC64;
```

Query Match 88.6%; Score 1421; DB 13; Length 301;  
Best Local Similarity 98.9%; Pred. No. 3.8e-126;  
Matches 266; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```
QY 33 MKTTPKKLRSSNVEOKTSSPOLSVLEDLERNAKKKAALDKIRAKATTPAGFETWRRELA 92
Db 33 VQITPKKLRSSNVEOKTSSPOLSVLEDLERNAKKKAALDKIRAKATTPAGFETWRRELA 92
QY 93 EEKPYFKQOLMSFVADERSRHTVYPADQVYSTEMCDIODYKVYVILGODPYHGNQAHG 152
Db 93 EEKPYFKQOLMSFVADERSRHTVYPADQVYSTEMCDIODYKVYVILGODPYHGNQAHG 152
QY 153 LCFSVQKRPVPPPSLVNITKELCTDIDGFKPHGHGDLDSGMAKQGVLLNNAVITVRAHOAN 212
Db 153 LCFSVQKRPVPPPSLVNITKELCTDIDGFKPHGHGDLDSGMAKQGVLLNNAVITVRAHOAN 212
```

```
QY 213 SHKRGWETFDAYIKWLSVNEGVFLLMGSYAAKKGATIDRKRNHVLQAVHPSPLSAH 272
Db 213 SHKRGWETFDAYIKWLSVNEGVFLLMGSYAAKKGATIDRKRNHVLQAVHPSPLSAH 272
QY 273 RGFGLGCKHFSKANGLLKLSGTEPIWMRAL 301
Db 273 RGFGLGCKHFSKANGLLKLSGTEPIWMRAL 301
```

```
RESULT 3
Q9JTW8 PRELIMINARY; PRT; 306 AA.
ID Q9JTW8;
AC Q9JTW8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE URACIL-DNA GLYCOSYLASE NUCLEAR ISOROM.
GN UNG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129SV;
RX MEDLINE=20330343; PubMed=10871356;
RA Nilsen H., Steinsbekk K.S., Otterlei M., Slupphaug G., Aas P.A.,
RA Krokan H.E.;
RT "Analysis of uracil-DNA glycosylases from the murine Ung gene reveals
RT differential expression in tissues and in embryonic development and a
RT subcellular sorting pattern that differs from the human homologues.";
RL Nucleic Acids Res. 28:2277-2285(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF174485; AAF76936.1; -.
DR EMBL; BC011039; AAH11039.1; -.
DR HSSP; P13051; IAKZ.
DR MGD; MGI:109352; Ung.
DR InterPro: IPR003249; Urac-DNA_glycosyl.
DR InterPro: IPR002043; U_DNA_glycosylase.
DR Pfam; PF00315; UNG; 1.
DR ProDom; PD001589; Urac_DNA_glycosyl; 1.
DR ProSITE; PS00130; U_DNA_GLYCOSYLASE; 1.
SQ SEQUENCE 306 AA; 33926 MW; CE2D5192936CE6EA CRC64;
```

Query Match 65.6%; Score 1051; DB 11; Length 306;  
Best Local Similarity 72.0%; Pred. No. 3.7e-91;  
Matches 188; Conservative 31; Mismatches 42; Indels 0; Gaps 0;

```
QY 41 RSSNVEOKTSSPOLSVLEDLERNAKKKAALDKIRAKATTPAGFETWRRELAEPKPYFK 100
Db 41 RSSNVEOKTSSPOLSVLEDLERNAKKKAALDKIRAKATTPAGFETWRRELAEPKPYFK 100
QY 101 QLMSFVADERSRHTVYPADQVYSTEMCDIODYKVYVILGODPYHGNQAHGLCFSVQKRP 160
Db 101 KLGMFVAEERNHNKHYVRRPQVFTWQMDIRQKVYVILGODPYHGNQAHGLCFSVQKRP 165
QY 161 VPPPSLVNITKELCTDIDGFKPHGHGDLDSGMAKQGVLLNNAVITVRAHOANSKDKGWE 220
Db 166 VPPPSLENIFKEHSTIDIDFVHGHGDLDSGMAKQGVLLNNAVITVRAHOANSKDKGWE 225
QY 221 TPTDAVTKWLSVNEGVFLLMGSYAAKKGATIDRKRNHVLQAVHPSPLSAHRGFLGCKR 280
Db 226 QFTDAVAVSNQNLNSGLVFLMGSYAOKKGVIDRKRNHVLQTAHDSPLSVHHRGFLGCRH 285
QY 281 FSKANGLLKLSGTEPIWMRAL 301
Db 286 FSKANGLLKLSGTEPIWMRAL 306
```

```

RESULT      4
P97285
ID          P97285      PRELIMINARY;      PRT;      306 AA.
AC          P97285;
DT          01-MAY-1997 (TREMBLrel. 03, Created)
DT          01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT          01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE          UTRAC1-DNA GLYCOSYLASE.
GN          UNG OR UNG2.
OS          Mus musculus (Mouse).
OC          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX          NCBI_TaxID=10090;
RN          [1]
RP          SEQUENCE FROM N.A.
RX          MEDLINE=97169285; PubMed=9016624;
RA          Nilsen H., Otterlei M., Haug T., Solum K., Nagelhus T.A., Skorpen F.,
RA          Krikan H.E.;
RT          "Nuclear and mitochondrial uracil-DNA glycosylases are generated by
RT          alternative splicing and transcription from different positions in
RT          the UNG gene."
RL          Nucleic Acids Res. 25:750-755(1997).
DR          EMBL; Y08975; CAAT0168.1; -.
DR          HSSP; P13051; 1AKZ.
DR          MGI; MGI:109352; Ung.
DR          InterPro; IPR003249; Utrac_DNA_glycosyl.
DR          InterPro; IPR002043; U_DNA_glycosylsc.
DR          Pfam; PF00315; UNG; 1.
DR          ProDom; PD001389; Utrac_DNA_glycosyl; 1.
DR          PROSITE; PS00130; U_DNA_GLYCOSYLASE; 1.
SQ          SEQUENCE      306 AA;      33952 MW;      CE2D519293C7FEB CRC64;

```

Query Match	65.28:	Score 1045:	DB 11:	Length 306:
Best Local Similarity	71.68:	Pred. No. 1.4e-90:		
Matches 187:	Conservative 32:	Mismatches 42:	Indels 0:	Gaps
QY	41	RSSNVEOKTSPOLSVEBLEEMAKNKKAALDKIKAKATPAGFGEFTWRELAAEPEKPYFK	100	
	:			
Db	46	KKARVEQWEGOSSPLSAEDLVITQNKKAALLRLAARNVPAGFESMKQQLCGEGKPYFV	105	
QY	101	QLMSEVADERSRHTVYPADQVYSSTEMCIDIQDVKVYILGODPYNHPGNOAHGLCFSYQKP	160	
	:			
Db	106	KLMGVADERNHKKVYPRPEQGVFWTQICSDIRDVKVYILGDDPYNHPGNOAHGLCFSYQRP	165	
QY	161	VPPPSLVNITKELCTDIDGCKRPHGHDLSGMAQGVLLNAVILTVRAHQNSHKDKGWE	222	
	:			
Db	166	VPPPSLENIKELSTDIDGCVNHPGHGDLSGMARQGVLLNAVILTVRAHQNSHKERKGE	225	
QY	221	TETPAVITKELSVNDEGVVFLMGSVYANKKATIDRKRHVLYQNVHPSPLSHRQFLCKH	280	
	:			
Db	226	QFTDVVSWMLQNSGLVFLFMGSLYQAKQGSVIDRKRHVLYQNVHPSPLSVYRKFGLCRH	285	
QY	281	FSKANGLEKLKSGTEPIWRAAL 301		
	:			
Db	286	FSKANELLQSGKKPIWKKEL 306		
RESULT	5			
Q93028		PRELIMINARY:	PRT:	313 AA.
AC	Q93028:	000719:	000637:	
DT	01-FEB-1997	(TREMBlrel. 02,	Created)	
DT	01-FEB-1997	(TREMBlrel. 02,	Last annotation update)	
DT	01-JUN-2001	(TREMBlrel. 17,	Last annotation update)	
DE	URACIL-DNA-GLYCOSYLASE.			
GN	UNG2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
NN	[1]			

```

RP SEQUENCE FROM N.A.
RX MEDLINE=97038682; PubMed=8884263;
RA Haug T., Skjopen F., Kvaloy K., Eftedal I., Lund H., Krokan H.E.;
RT "Human uracil-DNA glycosylase gene: sequence organization, methylation
RT pattern, and mapping to chromosome 12q23-q24.1.";
RL Genomics 36:408-416(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95010790; PubMed=7926048;
RA Haug T., Skjopen F., Lund H., Krokan H.E.;
RT "Structure of the gene for human uracil-DNA glycosylase and analysis
RT of the promoter function.";
RL FEBS Lett. 353:180-184(1994).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97169285; PubMed=9016624;
RA Nilsen H., Solum K., Haug T., Krokan H.E.;
RT "Nuclear and mitochondrial uracil-DNA glycosylases are generated by
RT alternative splicing and transcription from different positions in the
RT UNG gene.";
RL Nucleic Acids Res. 25:750-755(1997).
DR EMBL; X89398; CAA61578.1; -;
EMBL; Y09008; CAA70211.1; -;
DR HSSP; P13051; 1AKZ.
DR InterPro; IPR003249; Urac-DNA_glycosyl.
DR InterPro; IPR002043; U_DNA_glycylase.
DR Pfam; PF00315; UNG; 1.
DR ProDom; PD001589; Urac-DNA_glycosyl; 1.
DR PROSITE; PS00130; U_DNA_GLYCOSYLASE; 1.
SO SEQUENCE 313 AA; 34645 MW; A4B27E6198A9F9C0 CRC64;

```

Query Match	Similarity	63.5%	Score 1018.5	DB 4	Length 313
Best Local	Similarity	69.9%	Pred. No. 4.5e-88		
Matches 188	Conservative	31	Mismatches 47	Indels 3	Gaps 2
QY	35	ITPKFLRSSNVEOKT--SSPOLSEVBLEEMAKNKKAALDKIRAKATPAGFGETWRRELA	92		
DB	46	IPAKKAPAGQEEGRPPSP--LSAQDLRIQNKKAALLRLAANRPVGFSGSKKHLNSG	104		
QY	93	EEFEKYFKOLMSFVADENSRHTVYRPADQVYSTEKDIDQKVVLIGDPPYHNGQANG	152		
DB	105	EEGKPYFIKLMGFVAEERKHTVYRPPHQVFWTQMCIDIKVKVVLIGDPPYHNGQANG	164		
QY	153	LCFSVOKVRPPRPSLVNITYKELSTGIDCKHKGHDLSCMAKQVLLLAVALTVRANQAN	212		
DB	165	LCFSVQRPVPPPSLENIYKELSTGIDFVNHGHDLSCMAKQVLLLAVALTVRANQAN	224		
QY	213	SHKDGWETFTDAVYIKWLSVNREGVAFLLMSGYAHKKGATIRKRHHVLDQAVHPSLSAH	272		
DB	225	SHKEGWDGFTDAVYVSWLNQNSGLVFTLMGSIYAAKKGSADIRKRHHVLDQAVHPSLSY	284		
QY	273	RGFLCKHPSKANGILLKLGSTEPIMMAL	301		
DB	285	RGFFECRHFSKTNELLOKSGKKPIDMKEL	313		
RESULT	6				
Q99KR9					
ID	Q99KR9	PRELIMINARY;	PRT;	199 AA.	
AC	Q99KR9;				
DT	01-JUN-2001	(TREMBLrel. 17, Created)			
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)			
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)			
DE		SIMILAR TO URACIL-DNA GLYCOSYLASE.			
OS		Mus musculus (Mouse).			
OC		Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX		NCBI_TaxID=10090;			
FN		[1]			
RP		SEQUENCE FROM N.A.			
RC		TISSUE=MAMMARY TUMOR. WAP-TGF ALPHA MODEL. 7 MONTHS OLD. GROSS			
NC		TISSUE: ;			

RA Strausberg R.; to the EMBL/GenBank/DBJ databases.  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC004037; AAH04037.1; -.  
DR HSSP: P13051; 1AKZ.  
DR InterPro: IPR003249; Utrac\_DNA\_glycosyl.  
DR InterPro: IPR002043; U\_DNA\_glycosylase.  
DR Pfam: PF00315; UNG; 1.  
DR ProDom: PD001589; Utrac\_DNA\_glycosyl; 1.  
DR PROSITE: PS00130; U\_DNA\_GLYCOSYLASE; 1.  
SQ SEQUENCE 199 AA; 22526 MW; 264FAD2B1597AC4A CRC64;

Query Match 55.3%; Score 887; DB 11; Length 199;  
Best Local Similarity 78.9%; Pred. No. 6.7e-76;

Matches 157; Conservative 20; Mismatches 22; Indels 0; Gaps 0;

QY 103 MEFVADERSHRTYVPADQYVSTTEMCDIDQVKKVILGODPYHGPNOAHGLCFSVQRPV 162  
D 1 MGFVAERNNHKKYPPPEQYFTWTQMCDIRVKKVILGODPYHGPNOAHGLCFSVQRPV 60  
QY 163 PPSLVNITYELCTDIDGFKHPGHGDLGNAKQGVLLNVLTVRAHOANSHKDRGWET 222  
D 61 PPSLENIFFELSTDDIDGFVHPGHGDLGNAKQGVLLNVLTVRAHOANSHKDRGWEO 120  
QY 223 TDVAVIKMLSVNRGQVFLMGSAHKKGATIDRRKHVILQAVHPSPLSAHSGFLGCKHFS 282  
D 121 TDVAVSLMNLGSLVFLMGSTAKKGVYIDRRKHVILQAVHPSPLSVYRGFLGCRHFS 180  
QY 283 KANGLKLKSGTEPIPMNRAL 301  
D 181 KANELLQKSGKKPIPMKEL 199

RESULT 7  
ID Q9KPK8 PRELIMINARY; PRT; 226 AA.  
AC Q9KPK8;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE URACIL-DNA GLYCOSYLASE.  
GN VC2359.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
NCBI\_TaxID=666;  
RN NCB1\_TaxID=666;  
RX MEDLINE=20406833; PubMed=10952301;  
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umam L.A., Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D., Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragol I., Sellers P., McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O., Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C., Fraser C.M.;  
RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.";  
RL Nature 406:477-483(2000).  
DR EMBL: AE004306; AAF95502.1; -.  
DR HSSP: P12295; 3EUG.  
DR TIGR: VC2359; -.  
DR InterPro: IPR003249; Utrac\_DNA\_glycosyl.  
DR InterPro: IPR002043; U\_DNA\_glycosylase.  
DR Pfam: PF00315; UNG; 1.  
DR ProDom: PD001589; Utrac\_DNA\_glycosyl; 1.  
DR PROSITE: PS00130; U\_DNA\_GLYCOSYLASE; 1.  
KW Complete proteome.  
SQ SEQUENCE 226 AA; 25446 MW; 8C972727FFCD4EA4 CRC64;

Query Match 43.8%; Score 702.5; DB 16; Length 226;  
Best Local Similarity 61.1%; Pred. No. 2.2e-58;

Matches 132; Conservative 25; Mismatches 58; Indels 1; Gaps 1;  
QY 85 TWRELLAEFEKPYFKOLMSFVADER-SRHTVVPADQYVSTTEMCDIDQVKKVILGODP 143  
D 6 TMHVDVINEKQAVFOOTLQFESQROAGKVITPPADVAFNAFETFGVKKVILGODP 65  
QY 144 YHGPNOAHGLCFSVQKAVPPPSLVNITYELCTDIDGFKHPGHGDLGNAKQGVLLNVL 203  
D 66 YHGPNOAHGLCFSVLPCKVTPPSLVNITYELADIPGFOIPPHGYLOSMAGQGVLLNVL 125  
QY 204 LTVRAHOANSHKDRGWETFDVAVIKMLSVNRGQVFLMGSAHKKGATIDRRKHVILQ 263  
D 126 LTVEQGMAHSHANTGWETFDVAVDALNOHRNGLIFLMSKQAKQMDRQRHHVILMA 165  
QY 264 VHPSPSAHNRGFLGCKHFSKANGLKLKSGTEPIPMNR 299  
D 186 PHSPLSAHNRGFLGCKHFSKTNOLQAGIAPIMWQ 221

RESULT 8  
ID Q9JZAI PRELIMINARY; PRT; 219 AA.  
AC Q9JZAI;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE URACIL-DNA GLYCOSYLASE.  
GN NMB1222.  
OS Neisseria meningitidis (serogroup B).  
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
NCBI\_TaxID=491;  
RN NCB1\_TaxID=491;  
RX MEDLINE=20175755; PubMed=10710307;  
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E., Nelson W.C., Gwinn M.L., Deboy R., Peterson J.D., Hickey E.K., Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A., Mason T., Clecko A., Parksey D.S., Blair E., Clifton H., Clark E.B., Cotton M.D., Uterback T.R., Khouri H., Qin H., Vamathevan J., Gill J., Scarlato V., Maignani V., Plaza M., Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappelli R., Venter J.C.;  
RT "Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.";  
RL Science 287:1809-1815(2000).  
DR EMBL: AE002470; AAF41604.1; -.  
DR HSSP: P12295; 3EUG.  
DR TIGR: NMB1222; -.  
DR InterPro: IPR002106; AA\_tRNA\_ligase-II.  
DR InterPro: IPR003249; Utrac\_DNA\_glycosyl.  
DR InterPro: IPR002043; U\_DNA\_glycosylase.  
DR Pfam: PF00315; UNG; 1.  
DR ProDom: PD001589; Utrac\_DNA\_glycosyl; 1.  
DR PROSITE: PS00339; AA\_tRNA\_LIGASE-II-2; UNKNOWN\_1.  
KW Complete proteome.  
SQ SEQUENCE 219 AA; 24588 MW; 46DE58FA283B3F2 CRC64;

Query Match 40.5%; Score 648.5; DB 16; Length 219;  
Best Local Similarity 55.8%; Pred. No. 2.7e-53;  
Matches 121; Conservative 35; Mismatches 60; Indels 1; Gaps 1;

QY 84 ETWRRELLAEFEKPYFKOLMSFVADER-SRHTVVPADQYVSTTEMCDIDQVKKVILGODP 142  
D 2 DTNHADALGCKOORFOEILINAVROERLSQITTPRADVAFNAFRLAPRVKAVILGODP 61  
QY 143 YHGPNOAHGLCFSVQKAVPPPSLVNITYELCTDIDGFKHPGHGDLGNAKQGVLLNVL 202  
D 62 PYHAGCAQHAGLSAFVSGKIRIPPSLVNITYELTETDIEGFSIPAHGCLTAAEQGVLLNVL 121  
QY 203 VLTVAHOANSHKDRGWETFDVAVIKMLSVNRGQVFLMGSAHKKGATIDRRKHVILQ 262



Dh 122 VLVYRAGQASHALLGHERFTDVIQIOLATHRKHLVPMIMGVYAOQKGLIDSONHLILT 181  
QY 263 AVHPSPLSAHGRFLGCKHFSKANGLLKLGSTEPINMR 299  
Db 182 APHPSPLSAVRGFGCRHFSQANSYLSRHGIDPINMK 218

RESULT 9  
Q9RWH9 PRELIMINARY: PRT: 247 AA.  
AC 09RWH9;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DE URACIL-DNA N-GLYCOSYLASE.  
GN DR0689.  
OS Deinococcus radiodurans.  
OC Bacteria: Thermus/Deinococcus group: Deinococcales; Deinococcus.  
OX NCBI\_TaxID=1299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RI;  
RA MEDLINE=20036896; PubMed=10567266;  
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
RA Vamathavan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,  
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
RA Fraser C.M.;  
RT "Genome sequence of the radioresistant bacterium Deinococcus  
RT radiodurans RI.";  
RL Science 286:1571-1577(1999).  
DR EMBL; AE001926; AAF10269.1; -.  
DR HSSP; P13051; 1AK2.  
DR TIGR; DR0689; -.  
DR InterPro: IPR003249; UraC\_DNA\_glycosyl.  
DR InterPro: IPR002043; U\_DNA\_glycosylase.  
DR Pfam; PF00315; UNG; 1.  
DR ProDom; PD001589; UraC\_DNA\_glycosyl; 1.  
DR PROSITE; PS00130; U\_DNA\_GLYCOSYLASE; 1.  
KW Complete proteome.  
SQ SEQUENCE 247 AA; 27745 MW; B92EB3319AFA736 CRC64;

Query Match 39.6%; Score 635.5; DB 16; Length 247;  
Best Local Similarity 56.1%; Pred. No. 5.4e-52;  
Matches 124; Conservative 27; Mismatches 69; Indels 1; Gaps 1;

QY 79 PAQFGEMRELAAFEKPFKOLMSFVADERSHNYVPPADQVYSTECMDIODVYVI 138  
Db 20 PAULPEDWQALLPEFSAPYFHELTDFLRQERKEYTIYPPAPDVFNALRYTPLEGVKLI 79  
QY 139 IGGDDPYHGRQAAGLGFCSVQKVPVPPSLVNIYKELCTDIDGFGHPGGLSGMAKGV 198  
Db 80 LGDDPYHGRQAAGLGFCSVQKVPVPPSLVNIYKELCTDIDGFGHPGGLSGMAKGV 139  
QY 199 LINAVALTVRAHQANSRDKGMEFTDAVITKMLSVNRGVVFLMGSAVHKKGATIDRRH 258  
Db 140 LINAVALTVRAHQANSRDKGMEFTDAVITKMLSVNRGVVFLMGSAVHKKGATIDRRH 199  
QY 259 HVLQAVHPSPLSAHGRFLGCKHFSKANGLLKLGSTEPINMR 299  
Db 200 VYIESGHPSPLS-EQYFFGTGTPSPSKTNEALEKGRGPEWQ 239

RESULT 10  
Q9JUC4 PRELIMINARY: PRT: 219 AA.  
AC Q9JUC4;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE PUTATIVE URACIL-DNA GLYCOSYLASE (EC 3.2.2.).  
GN UNG OR NMA1384.  
OS Neisseria meningitidis (serogroup A).  
OC Bacteria: Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=65699;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=22491 / SEROGROUP A / SROTYPE 4A;  
RA MEDLINE=20222556; PubMed=10761919;  
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,  
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
RA Davies K.M., Davis P., Dellin K., Feltwell T., Hamlin N., Holtroyd S.,  
RA Jagers K., Leachter S., Moule S., Mungall K., Quail M.A.,  
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,  
RA Whitehead S., Spratt B.G., Barrall B.G.;  
RT "Complete DNA sequence of a serogroup A strain of Neisseria  
RT meningitidis 22491.";  
RL Nature 404:502-506(2000).  
DR EMBL; AL162755; CAB84630.1; -.  
DR HSSP; P12295; 3EUG.  
DR InterPro: IPR002106; AA\_tRNA\_ligase\_II.  
DR InterPro: IPR003249; UraC\_DNA\_glycosyl.  
DR InterPro: IPR002043; U\_DNA\_glycosylase.  
DR Pfam; PF00315; UNG; 1.  
DR ProDom; PD001589; UraC\_DNA\_glycosyl; 1.  
DR PROSITE; PS00339; AA\_tRNA\_LIGASE\_II\_2; UNKNOWN\_1.  
KW Hydrolase; Glycosidase; Complete proteome.  
SQ SEQUENCE 219 AA; 24685 MW; 46C576CC7E2C03F2 CRC64;

Query Match 39.6%; Score 634.5; DB 16; Length 219;  
Best Local Similarity 54.8%; Pred. No. 5.7e-52;  
Matches 119; Conservative 36; Mismatches 61; Indels 1; Gaps 1;

QY 84 ETWRELAAFEKPFKOLMSFVADERSHNYVPPADQVYSTECMDIODVYVI 142  
Db 2 DTWMDALGEGKQDPYFQELINAVRQERLSCGIIYPPADVFNARLTAFDKRVKAVILGOD 61  
QY 143 PYHGRQAAGLGFCSVQKVPVPPSLVNIYKELCTDIDGFGHPGGLSGMAKGV 202  
Db 62 PYHGRQAAGLGFCSVQKVPVPPSLVNIYKELCTDIDGFGHPGGLSGMAKGV 121  
QY 203 VLVYRAGQASHALLGHERFTDVIQIOLATHRKHLVPMIMGVYAOQKGLIDSONHLILT 262  
Db 122 VLVYRAGQASHALLGHERFTDVIQIOLATHRKHLVPMIMGVYAOQKGLIDSONHLILT 181  
QY 263 AVHPSPLSAHGRFLGCKHFSKANGLLKLGSTEPINMR 299  
Db 182 APHPSPLSAVRGFGCRHFSQANSYLSRHGIDPINMK 218

RESULT 11  
Q9U221 PRELIMINARY: PRT: 282 AA.  
AC Q9U221;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Y56A3A.29A PROTEIN.  
GN Y56A3A.29A.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Matthews L.;  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for

RT Investigating biology."  
 RL Science 282:2012-2018(1998).  
 DR EMBL: AL132860; CAB60520.1; -  
 DR HSSP: P13051; 1AKZ.  
 DR InterPro: IPR003249; UraC\_DNA\_glycosyl.  
 DR InterPro: IPR002043; U\_DNA\_glycosylase.  
 DR Pfam: PF00315; UNG; 1.  
 DR ProDom: PD001589; UraC\_DNA\_glycosyl; 1.  
 SQ SEQUENCE 282 AA; 31893 MW; A0814C43077E64BE CRC64;

Query Match 39.5%; Score 632.5; DB 5; Length 282;  
 Best Local Similarity 46.8%; Pred. No. 1.2e-51;  
 Matches 132; Conservative 46; Mismatches 93; Indels 11; Gaps 5;

QY 30 SKLKITPKKLRSSNVQKTSPPQLSVQLEERMAKKKALDKIRAKA-NPAGP----- 82  
 DB 2 SKTVRIPEMLKASAKSKRSASNTENIPKVPAGNENQVKKKLQAPETETILKSL 61  
 QY 83 -GETWRRELAEEFEKPYFKOLMSFVADERSR-HTVYPADQVYSTEMCDIDQKVVILG 140  
 DB 62 TGSWSKLEEEFKKGYISKIEKFLNSEVKNKGQVFPPIQITFMLLPFDELSVILG 121  
 QY 141 QDPYHGNQAHGLCFSVQKVPVPPPSLVNITYKELCTDIDGFKHPHGDLSGMAKQVLL 200  
 DB 122 QDPYHNDQAHGLSFSVQKGVKPPPSLKNITYKELESDIEGFKRPDHGNLGLWTFQGVML 181  
 QY 201 NAVLYRAHQANSKDRGWEFTDAVYIKMLSVNRE-GVFFLLGSSYAHKKGATIDRRKHH 259  
 DB 182 NAVLYRAHQANSKAKIGWTFETDVIYIRITISROSEKPIVFLWGGFAHKKKEELIDTKHV 241  
 QY 260 VLOAVHPSPLSAHGRFLGCKHFSKANGLKLSTGEPIINMAL 301  
 DB 242 VIKTAHPSPLSA-RKMWGCKCFKCNTELENSGRNPINMADL 282

RESULT 12  
 ID 092LU5 PRELIMINARY; PRT; 241 AA.  
 AC 092LU5;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DE PROBABLE URACIL-DNA GLYCOSYLASE PROTEIN (EC 3.2.2.2).  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 OX NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RA MEDLINE=21368234; PubMed=11474104;  
 RA Galibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,  
 RA Barloy-Hubler F., Barnett M.J., Becker A., Boisdard P., Bothe G.,  
 RA Bouty M., Bowser L., Burnmester J., Cadieu E., Capela D., Chain P.,  
 RA Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,  
 RA Gloux S., Godrie T., Goffeau A., Golding B., Gouzy J., Gujral M.,  
 RA Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,  
 RA Kahn M.L., Kalman S., Keating D.H., Kiss E., Komp C., Lelaure V.,  
 RA Masny D., Palm C., Peck M.C., Pohl T.M., Portetelle D., Purnelle B.,  
 RA Ramsperger U., Surzycki R., Thebault P., Vandewol M.,  
 RA Vorholter F.J., Weidner S., Wells S., Wong K., Yeh K.-C., Batut J.;  
 RT "The composite genome of the legume symbiont Sinorhizobium meliloti."  
 RL Science 293:668-672(2001).  
 DR EMBL: AL591792; CAC47505.1; -  
 KW Hydrolyase; Glycosidase; Complete proteome.  
 SQ SEQUENCE 241 AA; 27174 MW; 341ADE847ACE3802 CRC64;

Query Match 37.8%; Score 606; DB 16; Length 241;  
 Best Local Similarity 53.5%; Pred. No. 3.2e-49;  
 Matches 116; Conservative 32; Mismatches 67; Indels 2; Gaps 2;

QY 84 ETWRRELAEEFEKPYFKOLMSFVADERSR-HTVYPADQVYSTEMCDIDQKVVILG 142  
 DB 9 ESMKAVLGGFERHGYMAELKRLLEEKQGRQIRPGEVEFRALDPLDVRVILIG 68  
 QY 143 PYHGNQAHGLCFSVQKVPVPPPSLVNITYKELCTDIDGFKHPHGDLSGMAKQVLL 202  
 DB 69 PYHNDQAHGLCFSVRGVTRPPPSLVNITYKELESDL-GIPPARGFLESNARQGVLL 127  
 QY 203 VLYVRAHQANSKDRGWEFTDAVYIKMLSVNREGVFFLLGSSYAHKKGATIDRRKHH 262  
 DB 128 VLYVERGRASHQGRGERTDAVIRAVNDAQGVVIMMGSTYQRKAATVDRSHLVLT 187  
 QY 263 AVHPSPLSAHGRFLGCKHFSKANGLKLSTGEPIINMR 299  
 DB 188 AVHPSPLSAHAGFFGCRHFSKANAFLLSGIDPIDMR 224

RESULT 13  
 ID 09LIH6 PRELIMINARY; PRT; 330 AA.  
 AC 09LIH6;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DE URACIL-DNA GLYCOSYLASE-LIKE PROTEIN.  
 OS Arabidopsis thaliana (Mouse-ear cross).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COLUMBIA;  
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COLUMBIA;  
 RX MEDLINE=20363099; PubMed=10907853;  
 RA Nakamura Y.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.  
 RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,  
 RT TAC and BAC clones."  
 RL DNA Res. 7:217-221(2000).  
 DR EMBL: AP001303; BAB02221.1; -  
 DR HSSP: P13051; 1AKZ.  
 DR InterPro: IPR003249; UraC\_DNA\_glycosyl.  
 DR InterPro: IPR002043; U\_DNA\_glycosylase.  
 DR Pfam: PF00315; UNG; 1.  
 DR ProDom: PD001589; UraC\_DNA\_glycosyl; 1.  
 SQ SEQUENCE 330 AA; 36289 MW; 696BC91977309E19 CRC64;

Query Match 37.3%; Score 598.5; DB 10; Length 330;  
 Best Local Similarity 45.8%; Pred. No. 2.5e-48;  
 Matches 131; Conservative 36; Mismatches 92; Indels 27; Gaps 6;

QY 37 PKLRSSNVQKTSPPQLSVQLEERMAKKKALDKIR-----AKATPAG----- 81  
 DB 48 PRVYVTVTSVADDSG-LTPEQIARAENFKVAKSKRNLAVCSERYTKAKSEGCYPLS 105  
 QY 82 ---FGETWRRELAEEFEKPYFKOLMSF-----VADERSRHTVYPADQVYSTEMCDID 133  
 DB 106 ELVRESWMLKALPGEFHKPYAKSLSDLEIREITDSKS-PLIIPQHLINALNTTPFDR 164  
 QY 134 VKVYVILGQDPYHGNQAHGLCFSVQKVPVPPPSLVNITYKELCTDIDGFKHPHGDLSGMA 193  
 DB 165 VKTVIIGQDPYHGNQAHGLSFSVQKGVKPPPSLKNITYKELESDL-GCSPRRGNIQKWA 223  
 QY 194 KQGVLLNAVLYVRAHQANSKDRGWEFTDAVYIKMLSVNREGVFFLLGSSYAHKKGAT 253  
 DB 224 VQGVLLNAVLYVRSKOPNSHAKKGWEOFTDAVIQISTQKQEGVFLMGRIQAEKSKLI 283

QY 254 DRKHHVLAQVHPSPLSAHRGFLGCKHPSKANGLLKSTGEPINM 299  
 Db 284 DATKHHLTLTAHPGSLGSRNRCFPCRHRSRANQLLEENGIRPIDWQ 329

RESULT 14  
 Q9K320 PRELIMINARY; PRT; 227 AA.

AC Q9K320; 09K320;  
 DT 01-OCT-2000 (Tremblrel, 15, Created)  
 DT 01-OCT-2000 (Tremblrel, 15, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel, 19, Last annotation update)  
 DE URACIL-DNA GLYCOSYLASE (EC 3.2.2.-).  
 GN UNG.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriales;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Oliver K., Harris D.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA MEDLINE-97000351; PubMed-8843436;  
 RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,  
 RA Kinshi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 the 8 Mb Streptomyces coelicolor A3(2) chromosome";  
 RL Mol. Microbiol. 21:77-96(1996).  
 DR EMBL; AL359949; CAB95800.1; -;  
 DR HSSP; P13051; 1AKZ;  
 DR InterPro: IPR003249; Urag\_DNA\_glycosyl.  
 DR InterPro: IPR002043; U\_DNA\_glycosylase.  
 DR Pfam; PF00315; UNG; 1.  
 DR ProDom; PD001589; Urag\_DNA\_glycosyl; 1.  
 DR ProDom; PD00130; U\_DNA\_glycosylase; 1.  
 DR PROSITE; PS00130; U\_DNA\_GLYCOSYLASE; 1.  
 SQ SEQUENCE 227 AA; 25205 MW; 775BFB5A984C2982 CRC64;

Query Match 36.9%; Score 591; DB 2; Length 227;  
 Best Local Similarity 53.0%; Pred. No. 7.8e-48;  
 Matches 114; Conservative 35; Mismatches 64; Indels 2; Gaps 2;

QY 84 ETWRRELAEEFEKPYFKOLMSFVADERSRHTVPPADQVYSSSTEMCDIDQVYVILGODP 143  
 Db 9 ESKREVLGGLGLOOPYFKELMEFEVEERANGPYPPREVEFAALADANPPFRVKKVLVIGODP 68  
 QY 144 YHGPNQAGLCSFYOKPVPSPSLVNYIKELCTDIDGFGHGDLSGNAKOGVLLINAV 203  
 Db 69 YHGEGHGLCSFYVGVAVVPPSLRNYIKEMAEELD-TIPDNGYLPMAEGQVLLINAV 127  
 QY 204 LTVRAHQANSHKDRGNETTDAVITKLSVNRGAVVFLMGSYAHHKGGATIDRRKHHVLA 263  
 Db 128 LTVRAHQANSHKDRGNETTDAVITKLSVNRGAVVFLMGSYAHHKGGATIDRRKHHVLA 263  
 QY 264 VHPSPLSAHRGFLGCKHPSKANGLLKSTGEPINM 298  
 Db 188 AHPSPLSAKK-FFGSRPFTQINEAVAGOGHEPIDW 221

RESULT 15  
 Q9PA28 PRELIMINARY; PRT; 256 AA.  
 AC Q9PA28;  
 DT 01-OCT-2000 (Tremblrel, 15, Created)

DT 01-OCT-2000 (Tremblrel, 15, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel, 19, Last annotation update)  
 DE URACIL-DNA GLYCOSYLASE.  
 GN XE2692.  
 OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 OX Xylella.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-9A5C;  
 RX MEDLINE-2965717; PubMed-10910347;  
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
 RA Alvarenga R., Alves L.M.C., Araya J.E., Bata G.S., Baptista C.S.,  
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,  
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,  
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Hobeisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Matsukuma A.Y.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhani A., Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A., Jr., Pesquero J.B.,  
 RA Quaggo R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA da Silva A.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuchioka M.H.,  
 RA Valida H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Zelditch J., Zetzel J.C.;  
 RT "The genome sequence of the plant pathogen xylella fastidiosa";  
 RL Nature 406:151-159(2000).  
 DR EMBL; AEC04075; AAP85489.1; -;  
 DR HSSP; P12295; 3EUG;  
 DR InterPro: IPR003249; Urag\_DNA\_glycosyl.  
 DR InterPro: IPR002043; U\_DNA\_glycosylase.  
 DR Pfam; PF00315; UNG; 1.  
 DR ProDom; PD001589; Urag\_DNA\_glycosyl; 1.  
 DR ProDom; PD00130; U\_DNA\_glycosylase; 1.  
 SQ SEQUENCE 256 AA; 28187 MW; 4F21E7001BA1265B CRC64;

Query Match 36.9%; Score 591; DB 16; Length 256;  
 Best Local Similarity 53.2%; Pred. No. 9.2e-48;  
 Matches 116; Conservative 31; Mismatches 63; Indels 8; Gaps 3;

QY 85 TWRRELAEEFEKPYFKOLMSFVADERSRHTVPPADQVYSSSTEMCDIDQVYVILGODP 140  
 Db 24 SWRAHGNWMLRREMDLSAFL--RAKVAAGVSVYPPGQIPFAAEATPFQVKKAVILG 80  
 QY 141 QDPYHGNQAGLCSFYOKPVPSPSLVNYIKELCTDIDGFGHGDLSGNAKOGVLLINAV 200  
 Db 81 QDPYHGNQAGLCSFYVGVAVVPPSLRNYIKELCTDIDGFGHGDLSGNAKOGVLLINAV 139  
 QY 201 NAVLTVRAHQANSHKDRGNETTDAVITKLSVNRGAVVFLMGSYAHHKGGATIDRRKHHV 260  
 Db 140 NAVLTVRAHQANSHKDRGNETTDAVITKLSVNRGAVVFLMGSYAHHKGGATIDRRKHHV 260  
 QY 261 LQAVHPSPLSAHRGFLGCKHPSKANGLLKSTGEPINM 298  
 Db 200 LRAAHPSPLSAHRGFLGCKHPSKANGLLKSTGEPINM 237

Search completed: August 21, 2002, 10:28:33

Thu Aug 22 10:06:21 2002

us-03-758-017a-2.rspt

Job time: 235 sec